FUNCTIONAL ANALYSIS OF UDP-SUGAR: STEROL GLUCOSYLTRANSFERASES

VATSALA MALIK

A thesis submitted in partial fulfilment
of the requirements of the
University of Northumbria at Newcastle
for the degree of
Doctor of Philosophy

Research undertaken in the
School of Life Sciences
University of Northumbria

July 2011

Dedication

This work is dedicated to Mum and Dad, for without their support, courage and most of all, philosophy, this would not have been possible.

Abstract

Glycosyltransferases (GTs) are essential for the biosynthesis and diversification of many therapeutically important natural products. Of these, UDP-sugar: sterol glucosyltransferases (UGTs) (**2.4.1.173**) catalyse the synthesis of therapeutically important steryl glycosides (SGs). Guided by the sequence similarity with a previously characterised *N*-terminally truncated UGT from *Saccharomyces cerevisiae* (UGT51), this study reports the cloning of the gene fragment encoding the *C*-terminal catalytic domains from related yeasts and the expression and characterisation of their encoded products produced. *N*-terminally histidine tagged proteins were purified for *in vitro* assays against a panel of sterol and steroidal acceptors. Liquid chromatography-mass spectrometry (LC-MS) and kinetic analysis led to the successful characterisation of two novel UGTs from *Pichia angusta* and *Kluyveromyces lactis*. In addition, testosterone was shown to be utilized by all UGTs, including the previously characterised *S. cerevisiae* UGT51. Random mutagenesis of UGTs and homology modelling of the *S. cerevisiae* UGT revealed structural similarities with family 1 bacterial glycopeptide GTs. Given the structural and mechanistic similarities among GT family 1 UGTs, this approach may provide a template for genetic manipulation of novel UGTs from other members of the GT superfamily with a better understanding of catalytic domains and for broadening their scope in drug development. It may also aid the development of a generic process in the synthesis of SGs.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Term
µCi	micro curies
µM	micro molar
CAN	Acetonitrile
AGE	agarose gel electrophoresis
Amp	Ampicillin
APS	ammonium persulphate
BB	bromothymol blue
Bp	base pair
BSA	bovine serum albumin
CFE	cell-free extract
CFU	colony forming units
Ci	Curies
CR	cresol red
dATP	2’-deoxyadenosine 5’-triphosphate
dCTP	2’-deoxycytosine 5’-triphosphate
DE	directed evolution
dGTP	2’-deoxyguanosine 5’- triphosphate
Dh	*Debaromyces hansennii*
DhUGT51'	Dh r-protein
dNTPs	deoxynucleotide triphosphates
DPM	disintegrations per min
DTT	Dithiothreitol

dTTP	2’-deoxythymidine 5’- triphosphate
DVV	desvancosaminyl vancomycin
eP-PCR	error prone PCR
ESI	electron spray ionisation
EZC	PCR cloning of MMM using EZClone enzyme solution
EZK	PCR cloning of MMT using KOD DNA pol
FE	forced evolution
FPLC	fast performance liquid chromatography
GalT	Galactosyltransferase
GlcNAc	*N*-acetylglucosamine
Gtfs	glycopeptide group GTs from *A. orientalis*
HTS	high throughput screening
IPTG	isopropyl-beta-D-thiogalactopyranoside
Kan	Kanamycin
Kb	kilo bases
kDa	kilo Dalton
Kl	*Kluveromyces lactis*
KlUGT51'	Kl r-protein
Kp	*Kluveromyces polysporus*
KpUGT51'	Kp r-protein
LC	liquid chromatography
LIC	ligation independent cloning
LSC	liquid scintillation counting
ManT	mannosyltransferase
mCi	milli curies
mM	milli molar

MMM	mutant megaprimers using Mutazyme
MMT	mutant megaprimers using Taq
MS	mass spectrometry
MurG	*E. coli* UDP-*N*-acetylglucosaminyltransferase
OleD	oleandomycin GT from *S. antibioticus*
ORF	open reading frame
Pa	*Pichia angusta*
PaUGT51'	Pa r-protein
pDNA	plasmid DNA
Pg	*Pichia guilliermondi*
PH	pleckstrin homology
Pol	polymerase
Pp	*Pichia pastoris*
PpUGT51'	Pp r-protein
PR	phenol red
Prep	preparation
RM	random mutagenesis
r-plasmid	recombinant plasmid
r-protein	recombinant protein
Sc	*Saccharomyces cerevisiae*
ScUGT51'	Sc r-protein
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SG	steryl glycoside
SIM	selected ion monitoring
SpA	specific activity

TDP	thymidine diphosphate
TEMED	N,N,N',N'-Tetramethylethylenediamine
TG	testosterone glycoside
TLC	thin layer chromatography
TMHs	transmembrane helices
UDP	uridine diphosphate
UDPG	UDP-glucose
UDP*G*	UDP-[U-^{14}C] glucose
UDPGA	UDP-galactose
UGT	UDP glycosyltransferase
USGT	UDP-glucose : sterol glycosyltransferase
USS	UGT signature sequence
WT	wild type
X-gal	bromo-chloro-indolyl-galactopyranoside

Acknowledgements

Prof. Gary W. Black has not only provided me with the opportunity to undertake this project, but also has given invaluable and relentless support and guidance that went much beyond his primary role as the PI of this study. A word of acknowledgement would not suffice for his role in my overall development as an independent and aware academic researcher. Apart from expressing deep gratitude, I sincerely hope that I would be able to cherish his support and guidance for many more years to come.

I am deeply thankful to Dr. Lynn Dover who stepped in at a crucial stage of this project and provided guidance in the right direction. I would also like to acknowledge the following persons: Prof. Justin Perry for providing help and support whenever required, Dr. Melissa Brazier-Hicks (Durham University) has helped with radiochemical assays while Dr. Marcus Durrant has helped with the homology modelling work. I thank Dr. Tony Flinn and Dr. Julian Northen (Onyx Scientific) for providing funding as well as for useful advice.

I would like to thank Northumbria University for partly funding this project and for giving me the most forming years of my career so far, a journey that began as an undergraduate student in 2004.

I would like to extend an acknowledgement to all my colleagues, especially Lorenzo, Maria and Andrew for providing lighter moments in lab A321. Research technicians Edwin Ludkin provided critical support for LC-MS analysis, while Stephen Reed went out of the way to help in many instances. Dr. Meng Zhang has provided kind assistance with the MS identification of peptides.

A special word of thanks to Chris and Priti for being there, especially in the most trying times during the course of this project.

Finally, I owe my will to carry on to the continued love, faith and support of my family back in India, especially my brother Vaibhav for encouragement and grand dad for his blessings.

1. Introduction

1.1 Natural product glycodiversification

It has been estimated that 40-70% of all new chemical entities currently entering drug development programs possess insufficient aqueous solubility owing to their polycyclic and hydrophobic nature. This in turn leads to their poor oral bioavailability and therapeutic efficacy resulting from poor dissolution rate primarily in the gastrointestinal (GI) tract (Hauss 2007). In this regard, lipid-based formulations are well documented and proven to enhance drug absorption by direct delivery into systemic circulation by avoiding hepatic first-pass metabolism and by increasing membrane permeability. However, several factors influence the success of lipid-based formulations such as viscosity and consistency of dosage that can affect drug release *in vivo*. Furthermore, reproducing the complex GI environment *in vitro* represents a major challenge for drug dissolution studies along with additional concerns such as ageing which may affect the overall drug bioavailability (Wasan 2001).

Many new drugs under development are therefore poorly water soluble and have non-ideal pharmacokinetic properties. One strategy is the use of pro-drugs, where the drug substance is linked to a water soluble moiety such as a carbohydrate which increases solubility and potential drug availability (Luzhetskyy *et al.* 2007; Seibel *et al.* 2006). The drug is subsequently released by normal metabolic processes, ideally at the desired site of action. In a similar manner, drugs are commonly metabolised by conversion *in vivo* by specific host-encoded endogenous enzymes into soluble carbohydrate derivatives which can then be excreted (Luzhetskyy *et al.* 2005). For example, the systemic the detoxification

and removal of steroid hormones takes place by glucuronic acid conjugation (Meech Mackenzie 2010; Stevenson 1999).

The synthesis of appropriate carbohydrate derivatives can therefore be employed either as drug delivery tools or as a means to understand drug pharmacokinetics such as adsorption, distribution, metabolism and excretion (ADME) (Erb *et al*., 2009; Hauss *et al*., 2007).

Approximately 50% of prescription products sold in Europe and the US are either natural products or derivatives thereof (Newman *et al.* 2003). Secondary metabolites from plants and microorganisms represent diverse as well as promising candidates for the development of new medicines in addition to their agronomic relevance (Ahn *et al.* 2009; Gachon *et al.* 2005). For example flavonoids, compounds with anti-inflammatory, antioxidant and antitumour roles form a major part of medicinal research. Of these metabolites, the vital role of natural products harboring sugar moieties, predominantly glycoproteins and glycolipids is well documented in eubacteria, eukaryotes, archaebacteria and viruses (Zhang *et al.* 2007). The occurrence of a myriad of complex carbohydrates such as polysaccharides, glycoproteins and glycolipids govern a diverse range of defined cellular functions such as energy storage and cell wall structural integrity in the form of starch and cellulose respectively in plants and other processes such as cellular interactions, cell adhesion and immune defense (Klement *et al*., 2007; Park *et al.* 2009b).

Therefore, since drug-linked sugar moieties participate in interactions between drugs and their biological targets consequently increasing their pharmacokinetic properties (ADME) and promote drug solubility and stability, glycosylated

natural products are being exploited by the pharmaceutical industry as potential leads in drug discovery and development (Kohara *et al.* 2007). Newman *et al.* (2003) have extensively reviewed the prospects of glycosylated products as sources for new drugs. For example several glycosides produced by microorganisms such as anthracyclines, amino-glycosides, macrolides and plants such as lignans and anthraquinones have been used as biologically active natural products in the treatment of various diseases. Some examples of these uses have been summarized in **Table 1.1.** Biotech companies such as Neose (Horsham, PA) for example have therefore invested billions of dollars with a view to improve glycosylation of medically useful glycoproteins.

Glycosylated natural product	Mode of action	Pharmaceutical use
Glycosylated natural products from microorganisms		
anthracyclines (daunomycin, doxorubicin, aclacinomycin A, nogalamycin)	DNA intercalation	daunomycin- acute lymphoblastic/myeloblastic leukaemias; doxorubicin- breast cancer, juvenile tumours, soft tissue and aggressive sarcomas; aclacinomycin A, nogalamycin- tumours with a acquired resistance to daunomycin
Macrolides (erythromycin, oleandomycin, spiramycin, josamycin, midecamycin)	binding to 50s rRNA	respiratory tract and soft tissue infections caused by streptococci, pneumococci, staphylococci and enterococci
coumarins (novobiocin)	gyrase inhibition	infections caused by *Staphylococcus epidermidis*
glycopeptides antibiotics (vancomycin (*Amycolatopsis orientalis*), teicoplanin)	binding to N-acyl-D-Ala-D-Ala termini of non-crosslinked lipid-PP-disaccharide-pentapeptides	fatal infections by gram positive bacteria
Glycosylated natural products from plants		
anthraquinones	absorption, inhibition of Na^+/K^+-ATPase	laxative agents
flavonoids (hesperidin, rutin, vitexin, hyperosid, apigenin-glycosides, luteolin-glycosides)	radical-oxygen scavengers	hesperidin, rutin- tone improvement in vein walls; vitexin, hyperosid- treatment of congestive heart failure; apigenin-glycosides, luteolin-glycosides- improvement in blood circulation
etoposide	DNA-topoisomerase II inhibition	anticancer therapy

Table 1.1 Glycosylated plant and microorganism derived natural products and their pharmaceutical uses (Luzhetskyy *et al.* 2008)

Structure-based approach to drug design is at the forefront of the pharmaceutical industry and therefore structural modification of natural metabolites is an emerging research area. In addition, from a chemical standpoint, sugar conjugation results in both increased stability (through the protection of reactive nucleophilic groups) and water solubility (Luzhetskyy *et al*. 2008).

Understanding the specific roles that glycans play in the disease process could offer new targets for specific therapeutic intervention of these events (Dove 2001). For example highly modified glycans on tumour cell surfaces that mediate carbohydrate-protein and carbohydrate-carbohydrate interactions are involved in critical events in the progression, dissemination and invasion of cancer cells (Barchi 2000; Gorelik *et al.* 2001) as depicted in **Fig. 1.1.** Antitumour drug design strategies therefore encompass targeting the production or interaction of tumour-associated glycans by carbohydrate based drugs or vaccines that mimic the sugars involved in these interactions. For example cytotoxicity tests with pluramycins, antibiotics consisting of an aryl-*C*-glycoside core, have demonstrated potency against several murine and human cell lines (Bililign *et al.* 2005).

Understanding the biochemical implications of these altered cell surface glycans has experienced a renaissance in the past decade. For example cell-wall proteomic profiling and direct mass spectrometric (MS) analysis of fungal wall glycoproteins could help identify suitable vaccine candidates (Yin *et al.* 2008).

In Nature, diversification of carbohydrate structure and function occurs in two ways: (i) novel sugar structures are generated through enzymatic modification of the functional groups of a common sugar precursor- this process is most elaborate in prokaryotic cell wall and secondary metabolite biosynthesis (Klement *et al*.

2007); (ii) a variety of glycoforms with differing regio- or steriochemistries are created by the attachment of sugar substrates to 'acceptor' molecules that include natural product aglycones, proteins, lipids and other sugars by the action of specific glycosyltransferases (GTs) (Thibodeaux *et al.* 2007) For example glycoprotein based targets for drug design consist of oligosaccharide chains covalently linked to an amino acid backbone (*N*-linked or *O*-linked), the structural composition of which is defined by the expression of GTs and glycosidase genes in different cells and the availability of their sugar nucleotide cofactors ('donors') of a specific saccharide (Barchi 2000). This makes the prospects of studying GTs as potential drug targets very exciting.

In addition, despite their versatility and potential in drug development, there are concerns over using natural carbohydrates as drugs. This is because monovalent interactions between proteins and carbohydrates are relatively low compared to known drugs and their acceptors (Barchi 2000). Moreover, most oligosaccharides are highly water soluble and therefore prone to degradation by digestive glycosidases or clearance by hepatic receptors, highlighting concerns over overall drug bioavailability. Furthermore, chemical synthesis of oligosaccharides involves regio, stereochemistry of bond formation to be controlled for synthesis of well defined products (Erb *et al.* 2009). This presents as a laborious and challenging process that requires additional protection and deprotection steps to obtain the required stereochemistry of the final product, and proving industrially unfeasible owing to decreased product yield, waste generation and formation of reaction by-products (Palcic 1999; Robertson and Steer 2004). For example for the synthesis of the pharmaceutical compound quercetin **(Fig. 1.2)** containing five –OH groups,

each of which can form glycosidic bonds, four of these must be protected to synthesize a specific desired monoglycoside (Lim and Bowles 2004).

This Chapter will highlight the potential of GTs for the synthesis of glycosylated compounds. It will shed light on GTs as a means to understanding complex carbohydrate synthesis. In addition, the prospects for engineering GTs with a view to better understand catalytic domains will be discussed.

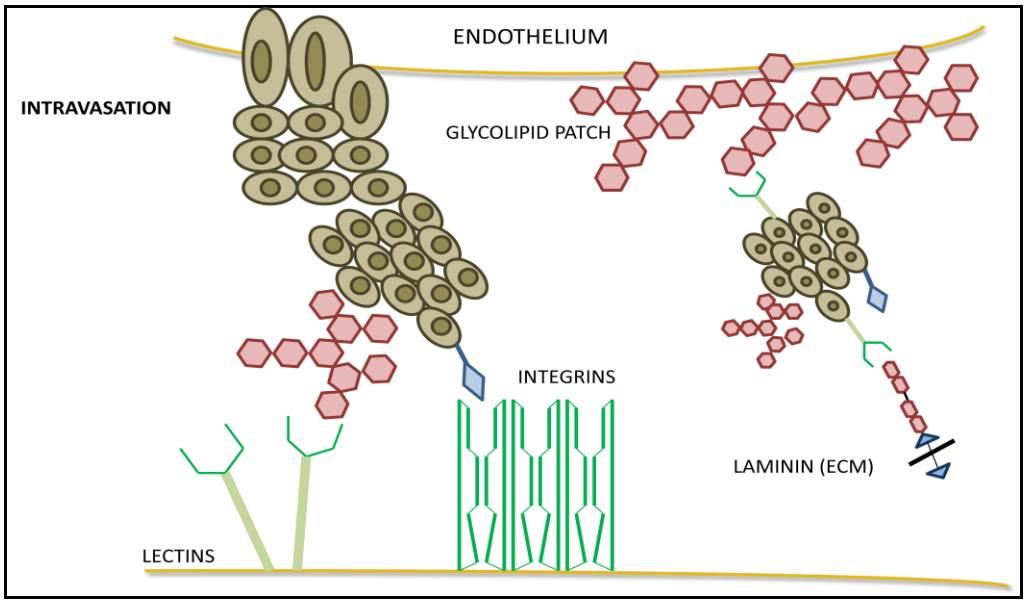


Fig. 1.1 Intravasation and metastasis of tumour cells by selective carbohydrate-mediated interactions (adapted from Barchi, 2000). Sugar chains (pink hexagons) bind with carbohydrate binding proteins (lectins) at surfaces of other cells- endothelial or tumour (brown ovals). These interactions are promoted by specific peptide sequences (blue arrowheads) recognised by cell adhesion molecules (integrins). Tumours interact with endothelial cell surface leading to intravasation. Binding with extracellular matrix (ECM) proteins (laminin) through polylactosamines (double hexagons) promotes release of proteases causing matrix degradation and tumour metastasis.

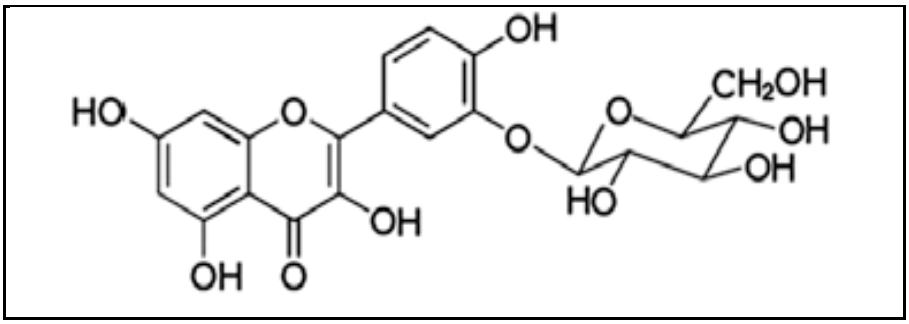


Fig. 1.2 Structure of quercetin 3'-*O*-glucoside (Wang *et al*., 2009).

1.2 Glycosyltransferases

Due to the nature of some cellular functions wherein carbohydrate moieties act as cellular language that depends on specific carbohydrate structures, the biosynthesis of complex sugars involves the action of a diverse range of selective GTs generating an infinite number of glycoconjugates and polysaccharides (Aharoni *et al.* 2006). GTs catalyze the sequential transfer of the monosaccharide units of their activated donor (usually nucleotide donor) to specific acceptor molecules such as lipids, saccharides and proteins resulting in the formation of a glycosidic bond (Hashimoto *et al.* 2009).

Activated sugar dependent transferases account for the vast majority of naturally occurring glycosyl transfer reactions and even though the range of acceptors is highly diverse, the donors are typically restricted to monosaccharides (glucose, galactose, mannose, xylose, fucose, arabinose etc) linked to a nucleoside mono- or di- phosphate leaving group (NMP or NDP) such as uridine diphosphate (UDP), GDP and CMP (Thorsoe *et al.* 2005). Of these, UDP-glycosyltransferases (UGTs) account for more than 60% of all GTs characterised so far (Hu Walker 2002).

1.2.1 GTs across genomes

Even though in eukaryotes, only ten different NDP sugars are utilised in the majority of biosynthetic glycosylation reactions, prokaryotic NDP-sugars are much more diverse (Thorson *et al.* 2001). Bacterial GTs therefore serve as excellent targets for drug development since most of these sugars are absent in humans. Since these sugars are important constituents of the bacterial cell wall,

their biosynthesis involves mechanistically novel enzymatic steps that have been actively targeted for inhibitor development. For example CHLαGcT, a cholesterol glucosyltransferase from a peptic and gastric ulcer causing bacterium *Helicobacter pylori* represents an excellent drug target due its vital role in the synthesis of cell wall lipids (Lebrun *et al.* 2006; Lee *et al.* 2006).

On the other hand, considering the vast number of glycosides in plants compared to other eukaryotes, it was believed that plant derived GTs would be more promiscuous. Even though plants GTs are known for their highly regio- and stereo selective nature, they can recognise common features on a range of substrates including hormones, secondary metabolites or xenobiotics. For example, glycosylation at the 3-OH position of certain plant pigments such as anthocyanins is crucial for the stability of the aromatic ring (Fukuchi-Mizutani *et al.* 2003; Gachon *et al.* 2005). This may imply that a wide range of substrates are accepted by plant GTs. For example Offen *et al.*, (2006) described the promiscuity of *Vv*GT1, a grape flavonoid UGT with respect to both the acceptor and donor substrates. It is important to note however that even though *in vitro* studies suggest plasticity in GT reactions, low substrate specificity could lead to undesired side reactions within the natural environment unless activity is regulated under transcriptional control (Osmani *et al.* 2009).

In addition to prokaryotes and other eukaryotes, glycosylation is a well established phenomenon in fungi where complex polysaccharides and glycolipids such as glucans and chitin form vital cell wall components (Gorelik *et al.* 2001; Hashimoto *et al.* 2009). Although glycolipids are known to contribute to the biophysical properties such as ion impermeability of membranes in these organisms, their specific roles have only recently being characterised. For

example membrane lipid- protein interactions are now known to govern the folding, localisation and function of membrane proteins. More recently Warnecke and Heinz (2010) have suggested an alternative but complementary approach to the current methods being employed for deciphering the function of specific glycolipids. Instead of creating deletion mutants of GTs involved in lipid metabolism that result in phenotypes with incomplete or lack of membrane systems in many cases, this approach would involve an exchange of a lipid GT in the organism by a heterologous GT with different specificity; for example substitution of a galactosyltransferase (GalT) with a glucosyltransferase. This would help in creating mutants without the complete loss of general functions but would highlight the specific role of the lacking membrane lipid.

N-linked and *O*-linked glycoprotein synthesis is however, is a well established process in yeasts such as *S. cerevisiae, P. pastoris* and *Candida albicans*, orchestrated by the action of specific GTs (Kim *et al.* 2006; Klutts *et al.* 2006; Mille *et al.* 2008). Golgi and ER GTs are involved in enhancing the stability and solubility of secretory proteins which also functions in the maintenance of fungal morphology in filamentous yeasts such as the *Aspergillus* species and are involved in virulence for example, in the fungal pathogen *Cryptococcus neoformans* that possesses a glycan rich cell wall surrounded by an intricate polysaccharide capsule (Goto 2007; Klutts *et al.* 2006). For example, an *A. nidulans* mutant lacking a protein *O*-mannosyltransferase (PMT), PmtC showed abnormal hyphae development and absence of conidia suggesting the vital role of PMTs in fungal development (Goto 2007).

The majority of GTs involved in eukaryotic cell wall polysaccharide and glycoprotein synthesis are Golgi-resident type II- membrane localised proteins

consisting of a short *N*-terminal cytoplasmic domain followed by a transmembrane domain, a stem region and the lumen facing *C*-terminal globular catalytic domain (Keegstra and Raikhel 2001; Qasba *et al.* 2008; Rabbani *et al.* 2006).

Quite recently, proteomic approaches coupled with MS have enabled the rapid and sensitive identification, as well as quantification of fungal wall glycoproteins and the involved GTs, which may pave the way to the development of novel diagnostic tools as well as vaccines (Lin *et al.* 2008; Yin *et al.* 2007).

1.2.2 Classification of GTs

The rapidly increasing number of genome projects has provided a wealth of sequence and biochemical information. GTs represent one of the most diverse groups of enzymes in biological systems and their classification thus provides a useful insight into their divergent evolution. Over the years, a few different systems of GT classification have been proposed.

It was initially believed that the most logical way to classify GTs would be according to their reaction mechanisms. Based on this, all GTs would be grouped as either ‘inverting’ or ‘retaining’ (Kapitonov and Yu 1999) (Reaction mechanisms described in section **1.2.2.1**).

GTs may also be classified based on differences in the carbohydrate glycosyl donors involved in the reaction (Compain and Martin 2001). Thus, they may either belong to the Leloir pathway, named after the 1970 Nobel Chemistry prize winner Luis Leloir following the discovery of uridine diphosphate D-glucose

(UDPG) and glycoside biosynthesis, or the Non-Leloir pathway. The former comprises of approximately 65% of all GTs, where donors may be activated by the addition of a mono- or diphosphate (CMP, GDP, UDP etc) forming a nucleotide sugar; the non-Leloir pathway GTs which utilise glycosyl phosphates or sucrose and other starch-derived oligosaccharides as donors (Weijers *et al*. 2008).

However, the universally recognised system for GT classification, based on amino acid sequence similarity (Campbell *et al.* 1997) is the Carbohydrate Active Enzymes database (CAZy) (http://www.cazy.org/) (Cantarel *et al.* 2009). This system comprises enzymes with roles in carbohydrate metabolism, biosynthesis and modification and is composed of 4 enzyme classes- glycoside hydrolases (GHs), glycosyl transferases (GTs), polysaccharide lyases (PLs) and carbohydrate esterases (CEs). Each class is further divided into families that incorporate both structural and mechanistic (stereochemistry of the glucoside linkage) features with members of one family assuming similar 3-D structural folds (GT-A, GT-B or GT-C) and catalytic mechanism (Liu and Mushegian, 2003; Luzhetskyy *et al.* 2005). However, sequence comparison cannot reliably predict catalytic mechanism (Osmani *et al.* 2009) and therefore, Coutinho *et al.* (2003) proposed sub-grouping of GTs into four clans as either inverting GT-A fold (clan I) or GT-B fold (Clan II) or retaining GT-A fold (clan III) or GT-B fold (clan IV).

So far, GTs have been classified into 92 CAZy families (designated as GTx) with over 50,000 open reading frames (ORFs) encoding putative GTs isolated from various plant, mammalian, fungal and prokaryotic sources.

This classification is underpinned by the assumption that signature structural motifs can be deciphered on the basis of primary protein sequence information (Henrissat and Davies 1997). Enzymes within a specific family may therefore have very different substrate specificities but be closely enough related to be considered divergent of one historical ancestor (i.e. homologous). It can also highlight convergent evolution whereby enzymes with very similar function and substrate specificity which would normally be grouped together by the IUBMB (International Union of Biochemistry and Molecular Biology) system of classification are shown to belong to several distinctly different families. For example, the large polyspecific families, such as GT2, consist of more than 3500 sequences that vary in donor and/or acceptor specificity and therefore demonstrate distinct functions (Breton *et al.* 2006). GT cclassification based on catalytic mechanism and structural fold will be further described in the following sub-sections.

1.2.2.1 Classification based on catalytic mechanism

GT catalysed reactions proceed by the sequential binding of the metal ion (if required for activity as a cofactor) and sugar nucleotide to the enzyme, followed by the acceptor (Qasba *et al.* 2005; Seibel *et al.* 2006) as shown in **Fig. 1.2 A**. The saccharide product is then ejected upon glycosyl transfer by 'inversion' or 'retention'. Analogous with glycosidases, GTs catalyse glycosidic bond formation with two possible stereochemical outcomes- inversion of the anomeric configuration (example UDP-glucose to β-glucoside) or retention of the anomeric configuration (example UDP-glucose to α-glucoside). There is limited knowledge

at the sequence level indicating differences between the two reaction mechanisms even though differences in UDP-binding amino acids have been suggested (Rosen *et al.* 2004).

Inverting GTs catalyse via a single displacement mechanism with nucleophilic attack by the acceptors at the C-1 (anomeric carbon) of the sugar donor (**Fig. 1.3 B**). This mechanism requires an active-site side chain to serve as a base to activate the sugar acceptor for nucleophilic attack by deprotonation because sugar hydroxyls themselves are poor nucleophiles (Charnock *et al.* 2001; Lairson *et al.* 2008). In addition, most GT Clan I enzymes characterised so far use a divalent cation (Mn^{+2} or Mg^{+2}) coordinated by the diphosphate moiety of the NDP-sugar. This ion binds with the conserved active site amino acid side chains to facilitate the departure of the NDP leaving group by stabilising the developing negative charge.

The inverting mechanism for GTs was first highlighted by the 3-D X-ray structure of both native and UDP-complexed SpsA, a GT-2 family member from *Bacillus subtilis* (Charnock and Davies 1999). Three side chain aspartate (D) residues are predominantly involved in binding the UDP-sugar. It showed that at the *N*-terminal, a nucleotide binding domain, D39 coordinated with the uracil base of the sugar donor whereas D98 and D99 (from a conserved DxD motif; described in section **1.2.2.2.**) was coordinated with ribose and UDP-bound Mn^{+2} respectively. In addition, the side chain carboxylate of D191 served as a base catalyst in the inverting mechanism.

Two reaction mechanism strategies have been proposed for retaining GTs. The earlier proposed strategy states the involvement of a double displacement

mechanism with the formation of a covalent intermediate. The S_N2 double-displacement mechanism involves a direct nucleophilic attack on the C1 of the UDP-sugar to form a β-linked glycosyl-enzyme intermediate. The leaving group phosphate itself plays the role of a base catalyst activating the incoming acceptor hydroxyl for nucleophilic attack (Lairson *et al.* 2008; Seibel *et al.* 2006).

Another more likely adopted strategy suggests that the enzyme uses a S_N1 transition state in which the approach of the attacking acceptor and the leaving donor are on the same side of the sugar ring (Zhang *et al.* 2007). It involves the production of a nucleophile by the dissociation of the leaving group NDP and attack on the transition state by the acceptor from the same side of the sugar ring as the scissile glycosidic linkage to the leaving NDP.

Although the both inverting and retaining catalytic mechanisms have been proposed, only the mechanism of inverting GTs has been well characterized so far (Qasba *et al.* 2005). This is due to the absence of any studies reporting the identification of the involvement of a catalytic nucleophile and/or a glycosyl-intermediate as has been suggested in the reaction mechanism.

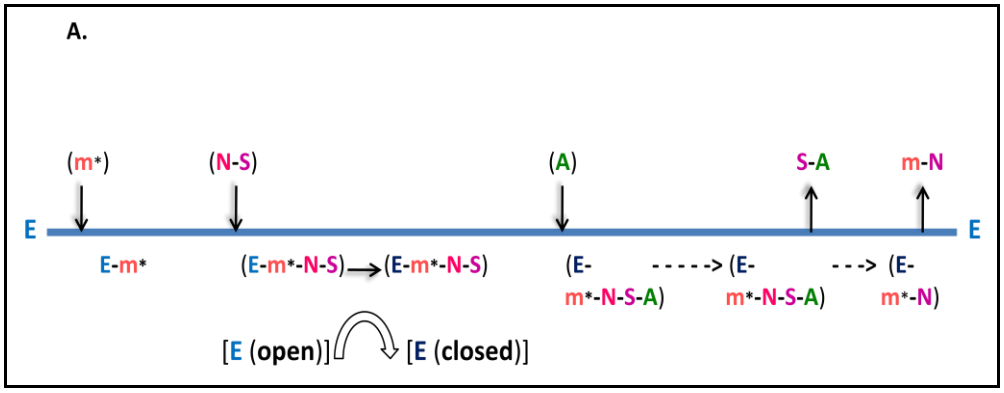


Fig. 1.3 A. Ordered reaction mechanism of GTs (Qasba *et al.* 2005). The binding of metal ion (m*) followed by the sugar-nucleotide donor (N-S) leads to a conformational change in the enzyme (E) creating the acceptor (A) binding site. Upon glycosyl (S) transfer to A, the product (S-A) is ejected followed by a reversal of E conformation to its native 'closed' state during which the metal-nucleotide complex (m-N) is ejected.

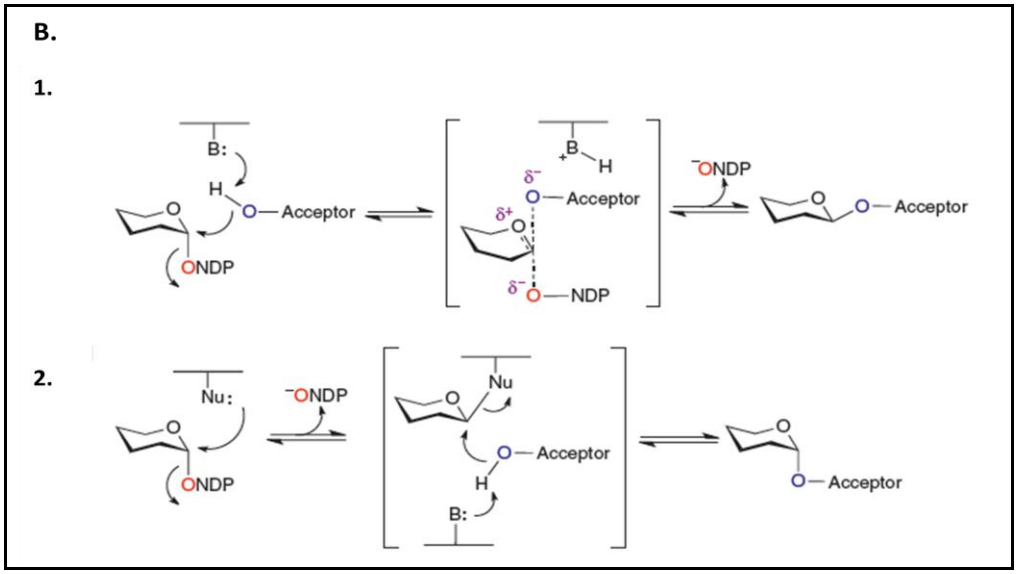


Fig. 1.3 B. Reaction mechanism of GTs (Thibodeaux *et al.* 2007). Inverting GTs-**1**.: general base (B)-catalysed S_N2 attack at C1 of the NDP-sugar donor by the acceptor forms an oxocarbenium-type transition state leading to the inversion of stereochemistry at the anomeric carbon; **2**.: The double displacement mechanism for retaining GTs involves the S_N2 attack of an active site nucleophile (Nu) at C1 of the NDP-sugar to form a covalent sugar–enzyme intermediate, followed by an S_N2 attack of the acceptor at C1 to form the glycosidic linkage with retention of anomeric configuration.

1.2.2.2 Classification based on structural fold topology

As compared with GHs where structural representatives from a large number of the 122 families are available, there is a relatively limited structural database for GTs (Hu and Walker 2002). Interestingly, in the former case, the highly diverse overall structural folds with conserved catalytic domains suggest evolutionary convergence during development of catalytic mechanisms. However, all NDP-sugar utilising GT structures so far fall into one of the two distinct structural fold families- the GT-A or GT-B. These observations indicate that a vast number of GTs may have evolved from a small number of progenitor sequences (Coutinho *et al.* 2003). Recently however, a third GT-C fold family (Liu and Mushegian, 2003) utilising lipid donors has also been described and is discussed in section **1.2.5.**

The two main GT folds (A and B) are well documented, with the overall topology of GT-A fold members consisting of two separate but closely associated β/α/β domains creating an active site in the centre. **Fig. 1.4 A.** depicts the structure of *Mycobacterium avium* GT (MAP2569c) that adopts this fold (Fulton *et al.* 2008). The *N*-terminal or NDP-sugar binding domain possesses a common DXD or EXD motif (Asp-X-Asp or Glu-X-Asp) flanked by non-polar residues. The DXD motif interacts with the phosphate group of the donor through the coordination the divalent metal ion cofactor. The acceptor binds at the *C*-terminal. Most Leloir pathway Golgi and endoplasmic reticulum residing GTs, as well as most prokaryotic GTs belong to this fold family (Hu and Walker 2002).

All GT-A fold members such as the α-mannosyltransferases (α-ManTs) and GalTs characterized so far require the presence of a metal ion cofactor that coordinates with the DXD side chain motif or its variants (example EDD, TDD) and was

therefore believed to be a characteristic feature of GT-A enzymes. However, one GT-A structure has been reported recently that does not possess this signature motif and Arg and Lys side chain residues were involved in stabilisation of electrostatic charges (Pak *et al.* 2006). Differences are also observed in the function of the DXD conserved motif residues in inverting and retaining GT-A enzymes. In the former case, both Asp residues interact with the metal ion whereas only the second Asp residue interacts with the metal ion in inverting members (Breton *et al.* 2006).

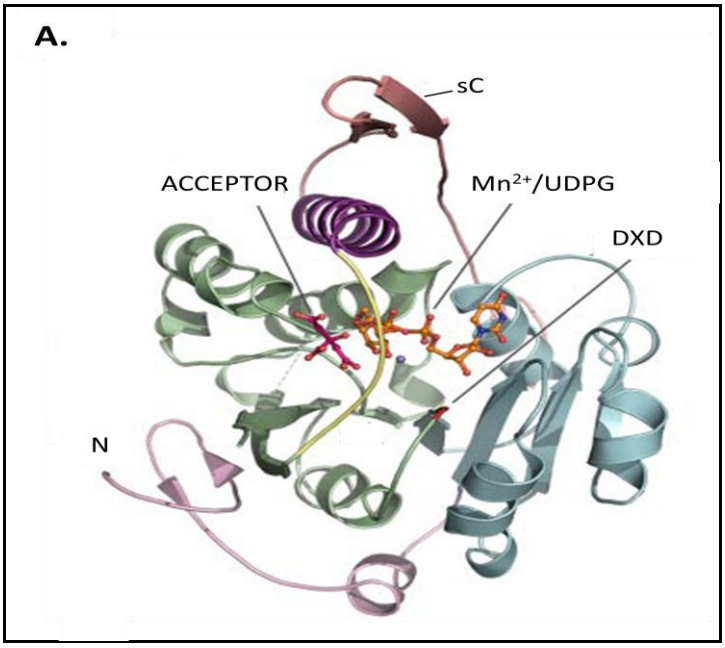
The second structural GT-B fold, was first studied in the T4 β-glucosyltransferase and members of this family consist of two separate Rossman-like β/α/β domains with a connecting linker region and a deep cleft between the domains which functions as the catalytic centre. **Fig. 1.4 B.** shows the GT-B fold of the crystal structure of a plant flavonoid GT (UGT71G1) (Shao *et al.* 2005). As opposed to the GT-A family, the *C*-terminal domain is the nucleotide binding domain in this case. The *N*-terminal domains are less conserved and are evolved to accommodate different acceptors (Kapitonov and Yu 1999). Members of this fold family comprise most of the prokaryotic GTs, such as those involved in glycosylation of natural secondary metabolites and cell wall synthesis (Thibodeaux *et al.* 2007) and are therefore more diverse compared to GT-A members.

As opposed to GT-A family enzymes, GT-B members catalyse in a metal ion independent manner (Breton *et al.* 2006). An example of this can be demonstrated by a study by Mulichak *et al*. (2003) where the crystal structure of Clan II GtfA from *Amycolatopsis orientalis* responsible for vancomycin biosynthesis showed that helix dipole and interactions with side chain hydroxyl and imidazole groups

were vital in stabilising the negative charge on the leaving group and therefore did not require the presence of divalent cations.

Iterative searches and multivariate data analysis employing several computational models and sequence information of characterized and putative GTs have revealed the prediction of a third GT-C fold (Liu and Mushegian 2003; Rosen *et al.* 2004; Wimmerová *et al.* 2003) now classified for 10 GT families including GT39, GT22 and GT51, GT66 and GT85 to name a few based upon two crystal structures. The predicted structure of the GT-C fold had indicated a large hydrophobic integral membrane protein which consisted of 8-13 transmembrane helices with a catalytic signature motif in the extracellular loop (Igura *et al.* 2008; Lairson *et al.* 2008). **Fig. 1.5 A.** shows the structure of the *C*-terminal domain of a GT66 oligosaccharyltransferase (OST) from *Campylobacter jejuni* (PglB) that adopts the GT-C fold (Maita *et al.* 2010). OSTs are involved in *N*-glycosylation by catalysing oligosaccharide transfer from lipid donors to polypeptide chains and are responsible for the virulence of pathogens such as *C. jejuni* (Maita *et al.* 2010).

So far all CAZy family members predicted to adopt this fold utilize lipid-phosphate activated donor sugar substrates (Lairson *et al.* 2008) as opposed to nucleotide-sugar donors or pyridoxal phosphate utilizing GT-A and GT-B fold enzymes. GT-C family members therefore belong to the Non-Leloir pathway of GTs. Thus, GT-C families mainly comprise of ManTs that use dolichol pyrophosphate-mannose as the sugar donor for the *O*-linked mannosylation of membrane lipids. In addition, GT85 arabinofuranosyltransferases (Aft) are involved in the synthesis of *M. tuberculosis* cell wall components as shown by Alderwick *et al*. (2006).



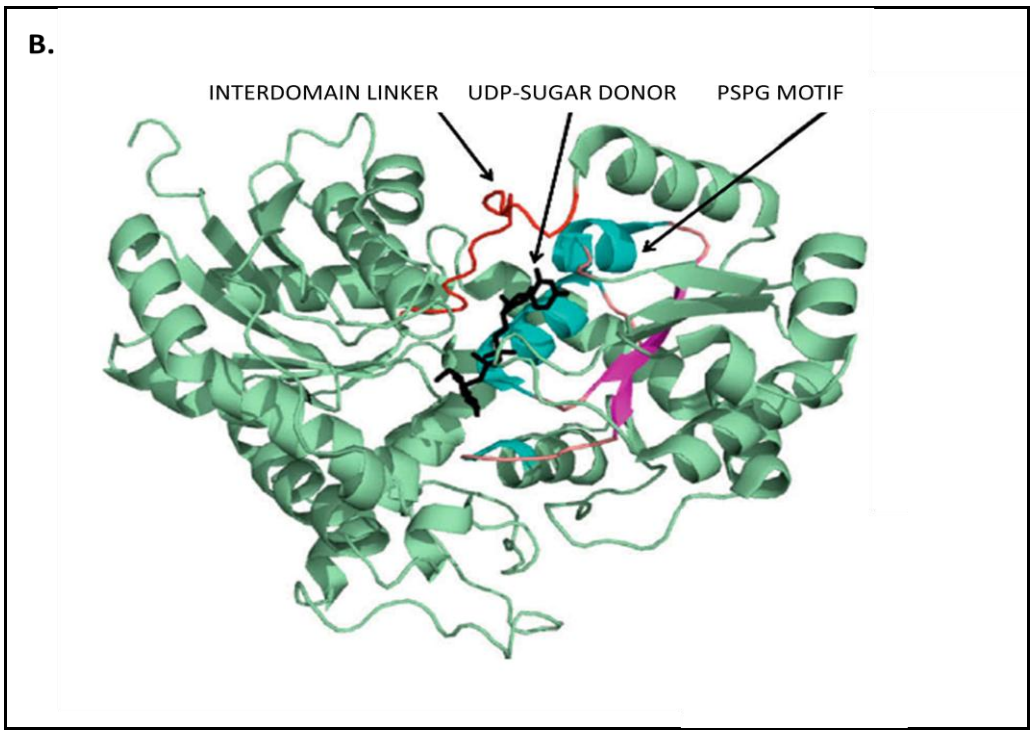
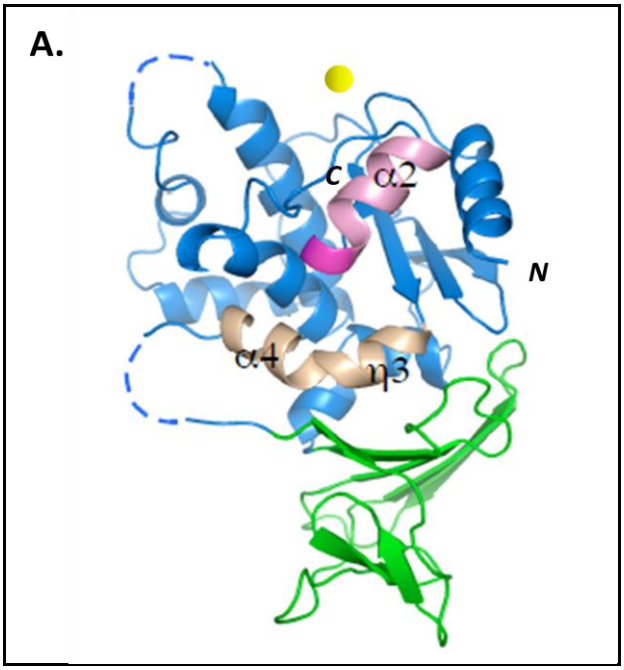


Fig. 1.4 Ribbon diagrams of GT structures depicting GT-A and GT-B folds. A: GT-A fold of the MAP2569c structure complexed with ligands (ball and stick representation) as indicated (Fulton *et al.* 2008). The *N*-terminal domain (pink), Rossman domain (cyan) and the catalytic DXD motif (red), acceptor binding site (green), flexible loop (yellow), helical region (purple) and the remainder of the *C*-terminal domain (salmon) are shown; **B**: GT-B fold of the UGT71G1 complexed with the UDP-sugar donor (ball and stick representation) as indicated (Osmani *et al.* 2009). The PSPG motif (cyan and pink) and inter-domain linker (red) are also shown.



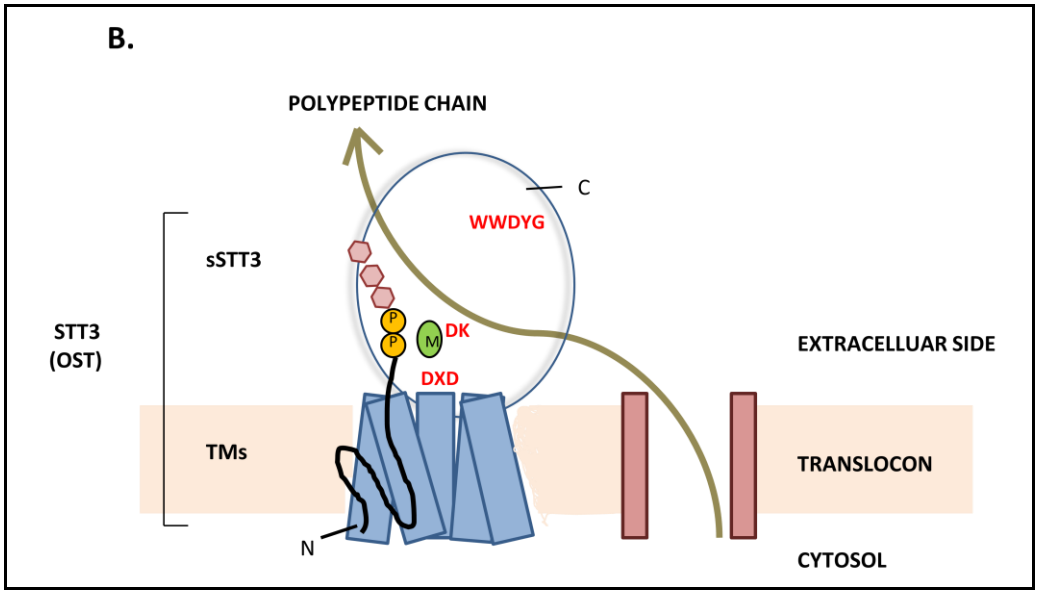


Fig. 1.5 GT-C fold of PglB highlighting the conserved residues of GT-C members. A. Structure of the *C*-terminal domain of PglB (Maita *et al.* 2010) showing a 10-stranded β-sheet structure (green) within the α-helical core domain (blue). The conserved WWDYG motif (magenta) is shown followed by a helix (pink). The long, kinked helix constituting the active site of OST is also shown (light brown). A bound metal cation is shown (yellow sphere) **B**: Schematic of the full length OST structure and possible reaction mechanism (adapted from Igura *et al.*, 2008). The *N*-terminal TMs and *C*-terminal globular domains are shown with the conserved motifs (shown in red, discussed in text) and bound cation (green) that are involved in *N*-linked glycosylation of the polypeptide chain using lipid-linked oligosaccharides (chain of pink hexagons and orange spheres).

1.2.3 GT Family 1 UGTs

The GT1 family currently consists of over 3,600 sequences with representatives from all phylae. Members of this family adopt the GT-B fold and mechanistically are inverting in nature, classifying them within Clan II according to the classification proposed by Coutinho *et al*. (2003). There is widespread interest in GT1 members as they are mainly recognised for the glycosylation of small molecules including important organic compounds such as terpenes, steroids, macrolides as well as are involved in detoxification pathways (Lairson *et al.* 2008; Lim and Bowles 2004). For example, the pollutant 3, 4-dicholoroaniline was detoxified by glucosylation using an *Arabidopsis* GT (*N*-GT) (Loutre *et al.* 2003).

Of the total 93 3-D structures of GTs solved so far, the GT1 family comprises 19, which is the maximum number of solved structures for any one GT family so far. This makes homology modelling an attractive tool for structure elucidation of novel GTs classified within this family.

GT1 includes UGTs as members where they represent the majority of all sequences within this family. Of these, plant derived sequences are most prominent. For example, approximately 120 *Arabidopsis* sequences are included within the UGT superfamily (Hansen *et al.* 2010). Sequences sharing at least 40% amino acid sequence identity are further classified within the same UGT subfamily (Osmani *et al.* 2009). All characterised as well as putative plant UGTs identified so far lack the presence of a signal peptide sequence or membrane spanning domains suggesting that these enzymes mainly function within the cytosol but may be associated as peripheral components of the endo-membrane system (Lim and Bowles 2004).

In addition, all GT1 UGT sequences, regardless of their origin share a conserved consensus sequence in the *C*-terminal domain denoted as the UGT signature sequence (USS) as originally proposed by Mackenzie *et al*. (1997). Among plant UGTs, this signature sequence is more conserved and is represented by a 44 amino acid PSPG (putative secondary product glycosyltransferase) box (Vogt and Jones 2000). This region has been characterised as the NDP-sugar binding domain (Lim and Bowles 2004; Mackenzie *et al*. 1997) and will be further described in the section **1.2.4.**

GT family 1 also includes a large number of glucuronosyltransferases (**EC 2.4.1.17**) involved in cellular homeostasis and phase II detoxification pathways (Offen *et al.* 2006). This class of GTs catalyse glucuronic acid transfer to potentially toxic lipophilic compounds such as bilirubin and steroid hormones thereby promoting their systemic clearance as highlighted earlier (Lim and Bowles 2004; Meech and Mackenzie 2010).

GT1 family UGTs contain a class of UDP-sugar: sterol glucosyltransferases (USGTs); also known as UDP-glucose: sterol 3-*O*-β-D-glucosyltransferases (**EC 2.4.1.173**). The first UGT of this class was purified from oat following which USGTs from both prokaryotes as well as eukaryotes were characterised (Idnurm *et al.* 2003; Kim *et al.* 2010; Madina *et al.* 2007). However, reports of sterols from bacteria are rare and only a few members of the *Mycoplasma* and *Helicobacter* species have shown to produce steryl glycosides from free sterols that they take up from their host organisms (Warnecke and Heinz 2010). Within the last decade, the identification and cloning of several sterol glycosyltransferase genes have aided the use of genetic approaches to investigate the current limited

understanding of the specific biological functions of steryl glycosides (SGs) in eukaryotes (Grille *et al.* 2010).

1.2.4 Conserved amino acid residues are involved in substrate recognition

Until recently, the molecular basis for GT donor and acceptor selectivity was unclear. However, in the past decade with a surge in the number of crystal structures that have become available from over 26 CAZy families within the GT 3D structure database (http://glyco3d.cermav.cnrs.fr/cgi-bin/rxgt/rxgt.cgi), it is now possible to better understand GT catalytic domains.

GTs are highly regio- and stereoselective with regards to substrate recognition *in vivo* (Weijers *et al.* 2008). However, as illustrated earlier in the case of plant GTs, modifications on the acceptor substrates can be tolerated *in vitro*. However, analysis of substrate specificity based on the contribution of individual functional groups for recognition and binding poses a great challenge due to the limited availability of unnatural oligosaccharides and high cost of NDP-sugar donors.

In the following sections (**1.2.4.1** and **1.2.4.2**), conserved amino acids and/or peptide motifs that play a major role in GT function will be discussed, with a special emphasis on GT-B fold UGTs as they are supported with better structural information.

1.2.4.1 Conserved residues in NDP-sugar recognition

A general pattern for substrate recognition can be observed among characterized GT-A and GT-B fold members. An inter-domain linker connects the *N* and *C*-terminal domains and the secondary structure elements are connected by less

conserved, flexible loop regions associated with substrate binding **(Fig. 1.4)**. For example, in the GT-A family, these flexible loops are associated with NDP-donor induced conformational change from open to closed where the loop acts as a lid covering the bound donor substrate creating an acceptor binding site (Qasba *et al.* 2005). This is also suggested to prevent hydrolysis of the sugar donor (Breton *et al.* 2006; Unligil and Rini 2000). The length and composition of the linker region may differ among different GTs and serves to define domain positioning and consequent substrate accommodation. For example, the inter-domain linker shows both length and sequence variability in the crystal structures of four GT1 family plant UGTs- *Mt*UGT71GT, *Vv*GT1, *Mt*UGT85H2 and *At*UGT72B1 (Osmani *et al.* 2009). This region differs significantly in comparison with crystal structures of bacterial GT1 members- GtfA, GtfB, GtfD from *A. orientalis* where it is much shorter and is directly involved in NDP-sugar binding (Mulichak *et al.* 2003; Mulichak *et al*. 2001; Mulichak *et al.* 2004). No single motif is conserved across GT families classified within the CAZy database (Hashimoto *et al.* 2009). However, the presence of a few conserved residues in the NDP-binding pocket across GT members from different species determine nucleotide sugar specificity and/or transfer to aglycones (Qasba *et al.* 2005). A classical example of this can be demonstrated by the presence of highly conserved His, Asp/Glu, Gln residues in the *C*-terminal PSPG motif (also known as the USS for UGTs as described earlier) of the GT1 family members. **Fig. 1.4 B.** indicates the position of the PSPG motif wherein apart from the other residues involved in binding the UDPG, the Asp, Gln and Trp residues of UGT71G1 form hydrogen bonds with the ribose group of the saccharide moiety of the donor (Osmani *et al.* 2009). A study by Potocka and Zimowski (2008) revealed that the Glu or Asp residue(s) in the

Solanum melongena (eggplant) UGT were crucial for UDPG binding whereas the His and Cys residues were important for both nucleotide and sterol binding.

The additional residues within the USS that do not directly interact with the sugar donor are believed to be involved in facilitating intramolecular interactions to stabilise the developing negative charge, as previously described (section **1.2.2.1**).

As earlier described, GT-A fold members coordinate the essential metal cation via the DXD motif. Although GT-B fold members lack this motif, they do exhibit a pattern of Gly and Ser/Thr residues located within the donor-binding domain (Klutts *et al.* 2006). These residues have been shown to interact with the phosphate group of the NDP-donor. For example the crystal structure of a GT-B family *N*-acetylglucosaminyltransferase (MurG) from *E. coli* involved in peptidoglycan biosynthesis contains three glycine rich loop motifs: one in the *C*-terminal that is suggested to bind the phosphate of the NDP-sugar and the other two within the *N*-terminal domain proposed to interact with the phosphate groups of the lipid-linked acceptor (Ha *et al.* 2000).

Although in the case of GT-B family UGTs, the majority of sugar substrate interactions occur in the structurally conserved *C*-terminal domain, some key *N*-terminal amino acids may also play a significant role in donor recognition and binding.

1.2.4.2 Conserved residues in acceptor recognition

In both the GT-A and GT-B members, the acceptor binding domain is generally less conserved suggesting that these may have evolved in order to accommodate a diverse range of aglycones (Hansen *et al.* 2009). However, some key residues

with charged side chains located within the acceptor binding pocket act as the general base triggering deprotonation of the acceptor as described in the reaction mechanism of GTs. In a majority of GT-B inverting enzymes, this base is His whereas the carboxylate group of the base catalyst Asp/Glu have been shown to be involved in deprotonation of the acceptor in some GT-A family members as illustrated earlier. The specific roles of these residues can be illustrated by the following examples. Mutation of the H19 and D117 resulted in complete loss of *At*UGT72B1 activity of the respective Ala variants, suggesting the vital role of these residues in the nucleophilic attack on the acceptor (Brazier-Hicks *et al.* 2007). It is also proposed that in a majority of inverting GTs, the histidine forms an Oacceptor-H-D triad where the Asp residue balances the charge on the His after proton abstraction. This was demonstrated by Shao *et al*., (2005) where substitution of the H22 and D121 with Ala in another plant UGT, UGT71G1 resulted in inactive variants suggesting that interaction of Asp with His was critical for catalytic activation by possibly forming an electron transfer chain.

1.2.5 Bioinformatic approaches for structure and function prediction of GTs

The number of families is continually increasing with the discovery and biochemical characterization of new GT genes. However, functional prediction of a putative GT based on sequence homology can be unreliable, as closely related sequences may express different catalytic activities. One such example can be presented by the UGTs of *Streptomyces fradie,* UrdGT1b and UrdGT1c that are responsible for the synthesis of the antibiotic urdamycin. These GTs differ in

acceptor specificity despite sharing 91% amino acid identity, (Hoffmeister *et al.* 2001) thereby exemplifying the complex nature of substrate specificity.

For large polyspecific families such as the GT2, it is especially difficult to predict the specificity. Sequence similarities are generally observed for the entire catalytic domain in the case of monospecific GT families, as opposed to polyspecific families where they are mostly restricted to only a portion of the catalytic domain (Breton *et al.* 2006). These sequences therefore cannot be classified functionally using bioinformatics (Hansen *et al.* 2010). Thus, novel putative GTs with no characterized homologues can only be identified on the basis of catalytic signature motifs employing bioinformatics strategies. This was further highlighted in the case of the GT-C family classification as demonstrated by the following examples.

As outlined earlier, bacterial GTs represent excellent potential drug targets and with the genome sequence of human pathogens such as *Mycobacterium tuberculosis* H37Rv available, cell wall synthesizing GTs for this pathogen for example are both an attractive and unexplored area. For example, the *M. tuberculosis* cell wall synthesizing enzymes EmbA, EmbB, and EmbC are currently one of the only few known drug targets for TB treatment and apart from *Mycobacterium* and related genera, they show no sequence similarity to homologues from other species (Berg *et al.* 2005).

Furthermore, GT-C members characterised so far use lipid donors and are pivotal in archaeabacterial and eubacterial cell wall synthesis, clearly suggesting the exclusivity of the involvement of these GTs in the process. However, GT-C classification will need to be further established on the basis of structural data as

so far, no characterized lipid phospho-sugar donor utilizing peptidoglycan synthesizing GTs (PGTs) have been assigned the GT-C fold. For example, crystal structures of the soluble domains two prokaryotic PGTs grouped under GT51 have recently been determined and have interestingly shown to adopt a novel λ-lysozyme-type fold even though lysozymes degrade peptidoglycan chains (Lovering *et al.* 2007; Yuan *et al.* 2007).

On the other hand, the first 3-D crystal structure of the GT-C assigned archaeal OST STT3 from *Pyrococcus furiosius* was determined recently (Igura *et al.* 2008). Similar to the *C. jejuni* OST **(Fig. 1.5 A)**, the *P. furiosius* OST adopts a novel α-helical fold surrounded by three β-sheet domains and is void of any characteristic Rossman-folds (Igura *et al.* 2008). Interestingly however, even though the *C*-terminal domain used for structural analysis consisted of the loop harbouring the predicted active site, the truncated protein resulted in a loss of activity suggesting the vital role of the *N*-terminal transmembrane region in OST activity. On comparison with the only other GT-C family solved structure an OST from the prokaryote *C. jejuni*, it was observed that the two OSTs shared significant structural similarity despite low sequence similarity (Maita *et al.* 2010) and are now grouped under GT66.

It is interesting to note that both the GT-C structures available so far have been determined using only the truncated soluble *C*-terminal domains of OSTs. These domains consist of conserved catalytic motifs in the first extracellular loop after the *N*-terminal domain (Berg *et al.* 2005) as shown in **Fig. 1.5 B**. The Asp in the conserved WWDYG motif acts as a catalytic base whereas the DK and DXD motifs are involved in binding lipid-linked donors through a transiently bound cation (Igura *et al.* 2008). Based upon this, a novel Aft from *M. tuberculosis*

(AftA) was identified and classified as a GT-C fold member (GT85) following structural sequence alignment with these conserved motifs (Alderwick *et al.* 2006).

The above studies therefore highlight the fact that computational methods based on motif-extraction, Hidden Markov models (HMM) prediction and structural comparison of sequenced genomes (Rosen *et al.* 2004; Wimmerová *et al.* 2003) may represent a more reliable approach in identifying novel GTs and predicting their structural folds as primary sequence properties may be more conserved despite low sequence similarity. However, determining acceptor specificity solely from the amino acid sequence may still present a challenge unless orthologs within a family are compared.

1.3 Diversifying functions of UGTs

The rational design of novel non-natural glycosylated compounds with desired properties requires a detailed knowledge of the function and mechanism of GTs. As illustrated in the previous sections, GT substrate specificity cannot be predicted from sequence comparison alone (Osmani *et al.* 2009). Even though the number of GT families classified in the CAZy database has almost doubled within the last decade, structural elucidation studies have struggled to keep pace with the increasing amount of sequence information of gene clusters involved in glycosylated natural product synthesis. Structure or sequence guided rational GT engineering attempts have largely been limited to single site-directed mutagenesis to decipher the roles of specific residues and/or elicit desired changes to enzyme properties (Hancock *et al.* 2006; Kim *et al.* 2010).

Since GTs display a two domain architecture with conserved 3-D structural folds, yet varied substrate (donor and/or acceptor) preferences, they represent ideal candidates for gene manipulation and engineering by methods such as directed evolution (DE) and combinatorial biosynthesis. These approaches would not only increase our present understanding of catalytic domains but would also help generate GTs with novel specificities (Hancock *et al.* 2006). Rational engineering and DE approaches are not entirely exclusive of one another because even though the latter approach does not necessarily require knowledge of the enzyme structure, knowledge-based design of GT variant libraries may improve overall engineering strategies (Dalby 2007).

The following sections will aim to review some of the approaches that have been employed to obtain novel enzymes with improved biochemical, physical and catalytic properties as well as desired substrate specificities.

1.3.1 GT metabolic pathway engineering of as a means to natural product glycodiversification

Due to the limited availability of NDP-sugars as well as the difficulty in accessing chemical derivatives using chemical synthesis routes, whole cell catalysis for the generation of glycosylated natural products is an evolving area. For example, in a recent study, a plant flavonoid GT was expressed in *S. cerevisiae* for the glycosylation of naringenin using endogenous UDPG (Werner and Morgan 2009).

The metabolic pathway of interest of a natural product producing strain can be manipulated to generate novel natural products through gene disruption and

heterologous expression of a gene encoding a GT catalysing the synthesis of the required glycosylated product. For example *dnrV*, the gene encoding 4-ketoreductase of the daunorubicin pathway of *Streptomyces peucitius* was disrupted and replaced by its homologue *eryBIV* that altered this biosynthetic pathway into producing epirubicin, an antitumour drug with improved therapeutic properties (Luzhetskyy and Bechthold 2008; Thibodeaux *et al*. 2007).

Combinatorial biosynthesis is the systematic modification and interchange of genes involved in the biosynthesis of natural products leading to the production of unnatural or hybrid natural products *in vivo* (Rix *et al.* 2002; Wohlgemuth 2005). In this emerging technology, whole cells are re-engineered by introducing genes/chimeras for carbohydrate-processing pathway enzymes from other organisms. So far, this approach has been mainly adopted for the biosynthetic pathways of the polyketide group of antibiotics such as erythromycin and vancomycin (Rix *et al*. 2002; Walsh *et al.* 2003) even though other examples have been reported. For example, in an attempt to produce a complex human *N*-glycan, the endogenous yeast glycosylation pathways of *P. pastoris* were replaced with chimeric GTs including mannosidases, *N*-acetylglucosaminyltransferases and the UDP-*N*-acetylglucosamine (GlcNAc) transporter (Hancock *et al.* 2006). However, in some cases, the low transformation efficiency coupled with conjugational incompetence of host strains may pose limitations to the size of the target gene/insert to be expressed (Hoffmeister *et al.* 2001).

1.3.2 DE of GTs

DE techniques mimic natural evolution processes such as random mutagenesis and sexual recombination (Dalby 2007). It is the generation and selection of a molecular library of enzymes with sufficient diversity for the altered function to be represented and evolved variants are selected and used as templates for the next round of evolution (Cohen *et al.* 2001; Rubin-Pitel and Zhao 2006).

In DE studies, gene fragments of interest can be engineered *in vitro* or in their natural cellular environments for glycosylated product synthesis. However, in the case of GTs, loss of enzyme activity does not result in a phenotypic change and therefore cannot form the basis of a cell-based selection method (Love *et al.* 2006).

1.3.2.1 In vitro DNA recombination and domain swapping

GT engineering and domain swapping studies have guided the identification of specific side chain residues or peptide motifs that govern substrate specificity in addition to offering a greater understanding of intra- and inter domain interactions (Osmani *et al.* 2009). However, due to the complexity of substrate recognition where several regions are involved, chimeric GTs have been generated using UGTs (Cartwright *et al.* 2008; Krauth *et al.* 2009). For example, based on previous knowledge of the specific protein region of 31 amino acids of UrdGT1b and UrdGT1c that defines both donor and acceptor recognition, Hoffmeister *et al.* (2002) created a chimeric gene library harbouring codons for the 10 non-identical residues in this region from both parental enzymes. In addition to exhibiting parental specificities for urdamycin synthesis, several variants displayed novel

glycosylation patterns and exclusively catalysed the synthesis of a derivative, urdamycin P that contained a branched saccharide appendage.

However, recently, Hansen *et al*. (2009) successfully applied the domain-swapping strategy to closely as well as distantly related family 1 plant UGTs. The chimeras were constructed based on sequence alignments and molecular modelling with known UGT structures and generated functionally active UGTs. For example chimeras constructed with *At*UGT71C1 and *At*UGT71C2 showed a three-fold increase in K_{cat} of etoposide glycosylation compared with the wild type enzymes. Increased catalytic efficiency towards etoposide and a novel regiospecificity towards *trans*-resveratrol was also observed for chimeras constructed with *At*UGT71C1 and *Sr*UGT71E1 even though these sequences displayed only 38% amino acid sequence identity.

1.3.2.2 DE by random mutagenesis

In random mutagenesis (RM), libraries containing random combinations of single and multiple point mutations across entire gene/gene fragment can be generated by techniques such as error-prone PCR (ep-PCR) using a low-fidelity DNA polymerase. Williams and co-workers (2007, 2008) evolved GTs involved in the glycosylation of specific macrolide antibiotics by random mutagenesis using ep-PCR. In one such example, a triple mutant (P67T/S132F/A242V) of the oleandomycin GT (OleD) from *Streptomyces antibioticus* showed a ~30 fold higher specific activity value against the fluorescent acceptor 4-methylumbelliferone (4-Mu) as compared to wild-type OleD (Williams *et al.* 2007). Following this, another triple mutant P67T/I112K/A242V of OleD was created with a 150-fold improved specific activity towards the nonnatural acceptor

novobiocic acid. It was found that Pro67 present in the N3 loop region and Ile112 were associated with substrate specificity and acceptor binding respectively. These mutants consisted of a polar/charged residue at position 112 and hydrophobic residue at position 242 which were found to contribute to enhanced activity (Williams *et al*., 2008).

1.3.2.3 Challenges for the DE of GTs

A hurdle to the scope of DE experiments involving enzymatic and chemoenzymatic *in vitro* synthesis of glycoforms is created by the limited availability of NDP-sugar donors. However, recent work has suggested improvements in chemical methods of NDP-sugar synthesis, as well as enzyme-based approaches for the synthesis and regeneration of sugar nucleotides (Mao *et al.* 2006; Masada *et al.* 2007; Pesnot and Wagner 2008; Ruffing *et al.* 2006). One such method for glycodiversification is glycorandomization which uses the inherent or engineered promiscuity of anomeric kinases and nucleotidyltransferases for the *in vitro* synthesis of accessible NDP-sugar donor libraries (Williams *et al.* 2007). However, the feasibility of the use of such methods requires further validation.

Also, most DE and genetic engineering platforms require an efficient screening system for evolved variants. However, due to the magnitude of the library generated (10^3-10^7 variants); the exploitation of RM studies is solely dependent on the availability of a robust HTS methodology based on a simple selection for glycosidic bond formation. Currently, for *in vitro* evolution studies, the lack of suitable and efficient high-throughput screening systems to select for the desired activity is well documented and represents the main focus of research for

expanding the scope of GTs. The majority of the current screening methodologies either involve linking the aglycone to a surrogate fluorescent acceptor (Trubetskoy and Shaw 1999) or derivitization of the acceptor (Cowan *et al.* 2008; Warnecke *et al.* 1999), which in both cases, leads to a detectable change linked with product formation. A few recent studies have however, reported the development of colorimetric screens for the rapid screening of GT activity (Deng and Chen 2004), the wider scope of which needs to be further assessed.

1.4 Aims and objectives

Glycosylation of drugs, which includes a major class of glycosylated steroids, significantly improves their stability, overall bioavailability and pharmacokinetic properties. However, the chemical synthesis of such compounds is a laborious and industrially unfeasible process and as such, paves the way for molecular based approaches for novel glycosylated product synthesis.

Family 1 UGTs comprise of a class of USGTs and can therefore be exploited for the biosynthesis of steroidal glycosides. The major aim of this study was to isolate and characterize the function of putative eukaryotic USGTs for the synthesis of glycosides from steroids and steroid derivatives. Warnecke *et al*. (1999) have experimentally shown that the majority of the *N*-terminal domain of ScUGT51 was not required for USGT activity. In addition to retaining its functional activity, the *N*-terminally truncated gene product was soluble and functionally stable when expressed in *E. coli*. The shortest *C*-terminal domain of ScUGT51 vital for enzyme function consists of sequence of 476 amino acids resulting from a deletion of the first 722 *N*-terminal amino acids (Warnecke *et al.* 1999).

This study has attempted to exploit the observation that ScUGT51' and related UGTs with *N*-terminal deletions could be available to detailed biochemical analysis as pure enzymes against a select panel of steroidal acceptors.

Furthermore, at present no structure is available for USGTs which limits their detailed functional analysis. This study also aimed to study the function of the isolated USGTs by employing DE studies by RM.

2. Materials and Methods

The following table (**Table 2.1**) lists manufacturers (and the abbreviated form) that were used to source chemicals, reagents and enzymes used within this study.

Manufacturer	**Abbreviation**
Acros	ACR
Alfa Aesar	AE
Amersham Biosciences	AB
Bioline	BL
Bio-Rad	BR
Brunswick Scientific	BRS
Fisher	FIS
Invitrogen	INV
Melford Laboratories Ltd.	MEL
MWG Biotech AG	MWG
National Diagnostics	ND
New England Biolabs	NEB
Novagen	NOV
Oxoid	OX
Perkin Elmer	PE
Sigma	SIG
Start Labs	SL
Stratagene	STR

Table 2.1 Manufacturers and their abbreviations used within the thesis.

The following Materials section (**2.1**) details the recipes of reagents and media used throughout this study. Protocols used for DNA extraction, cloning and protein expression as well as UGT assay protocols are detailed in the Methods section **2.2**.

2.1 Materials

2.1.1 Equipment

All general use equipment is listed in Appendix **D.**

2.1.2 Media

Unless otherwise stated all solutions were made up with 18.2 MΩ/cm H_2O purified by a Millipore Direct-Q® water purification system and were stored at room temperature. All media was prepared using distilled H_2O and sterilised by autoclaving at 121°C. Liquid media was stored at room temperature, and solid media at 4°C. All adjustments to the pH of solutions were achieved using 1 M HCl or 1 M NaOH unless stated otherwise.

2.1.2.1 Liquid media

LB (Luria-Bertani) broth

Component	Amount (g)/L distilled H_2O
Tryptone (OX)	10
Yeast extract (OX)	5
NaCl (SIG)	10

Components were sequentially dissolved and adjusted to pH 7.0 followed by autoclave sterilisation.

Low salt LB broth

Used for the preparation of electrocompetent cells for transformation. 1 g low salt LB (SIG) was dissolved in 50 ml LB media in a 250 ml conical flask and autoclave sterilised.

Liquid media for growing yeast cultures

All liquid media was autoclaved at 121 ^{o}C prior to use. For the growth of yeast and bacterial strains on solid media, plates were prepared with the addition of 2% w/v Agar (OX).

(a) YEPD broth

Component	Amount (g)/L distilled H_2O
Bacto-Peptone (BD)	20
Yeast extract (OX)	10
Glucose (SIG)	20

(b) YM Broth

Component	Amount (g)/L distilled H_2O
Yeast Extract	3
Malt Extract (OX)	3
Peptone (OX)	5
Glucose (SIG)	10

(c) 393/YPD Broth (pH 6.5)

Component	Amount (g)/L distilled H_2O
Peptone	20
Yeast extract	10
Glucose	20

(d) Broth 186

Component	Amount (g)/L distilled H_2O
Yeast extract	3
Malt extract	3
Glucose	10
Peptone (soybean)	5

2.1.2.2 Solid media

LB agar

For the preparation of agar plates 2% (w/v) agar (bacteriological agar No 1) (OX) was added to sterile LB broth. Agar was then made soluble by autoclaving the media and poured into Petri dishes upon cooling to 50°C.

2.1.2.3 Selective media

Antibiotics, IPTG, X-Gal

Antibiotics were made up at stock concentrations in sterile 18.2 MΩ/cm H_2O unless stated otherwise and were diluted to the appropriate concentration in media **(Table 2.2)**.

Antibiotic/IPTG	Stock concentration	Dilution of stock	Final concentration in media	Storage of stocks
Kanamycin (SIG)	10 mg/ml	1:100	100 μg/ml	4°C for up to 3 days
Chloramphenicol (SIG)	34 mg/ml in 100% (v/v) ethanol	1:1000	34 μg/ml	-20°C
Streptomycin (SIG)	10 mg/ml	0.75:1000	75 μg/ml	-20°C
Ampicillin	10 mg/ml	1:100	100 μg/ml	4°C
Isopropyl-beta-D-thiogalactopyranoside (IPTG) (SIG)	24 mg/ml	1:100	240 μg/ml	4°C
X-Gal	40 mg/ml in 100% (v/v) dimethyl formamide	1:500	80 μg/ml	-20°C

Table 2.2 Stock concentrations and final concentrations of antibiotics in media

For the addition of antibiotics to molten agar the media was cooled to ~55°C.

2.1.3 AGE -buffers, dyes and reagents

2.1.3.1 TAE running buffer (50x) (TRIS-HCl (2 M), Glacial acetic acid (1 M), EDTA (0.05 M))

Component	per L 18.2 MΩ/cm H_2O
TRIS (SIG)	242 g
Glacial Acetic acid (17.51 M) (FIS)	57.1 ml
EDTA (SIG) (0.5 M) solution pH 8.0	100 ml

Running buffer was diluted down to 1x with distilled H_2O prior to use for the preparation of agarose gels and use as reservoir buffer in gel tanks.

2.1.3.2 Ethidium bromide (EtBr) (SIG)

EtBr (10 mg/ml) was used diluted with 18.2 MΩ/cm H_2O to a working concentration of 10 μg / ml prior to use.

2.1.3.3 Sample loading buffer

Bromophenol blue (6x) sample loading buffer

Component	Amount
Bromophenol blue (SIG)	0.25% w/v
Glycerol (FIS)	30% v/v

DNA loading buffer without bromophenol blue (6x) (for DNA bands < 1Kb)

Component	Amount
Xylene cyanol (SIG)	0.25% w/v
Glycerol (FIS)	30% v/v

Loading buffer was diluted down to 1x with the sample prior to loading on an agarose gel.

2.1.4 General buffers and reagents and size standards

2.1.4.1 TE buffer (TRIS-HCl 10 mM, EDTA 1 mM pH 8.0)

Component	Volume (ml)/ L 18.2 MΩ/cm H_2O
TRIS (SIG)-HCl (0.5 M) pH 7.5	20
EDTA (0.5 M) pH 8.0 solution	2

2.1.4.2 STET Buffer

Component	Amount/volume/L
Sucrose	80 g
Triton X-100	50 ml
0.5 mM EDTA (pH 8.0)	100 ml
TRIS-HCl	6.06 g

2.1.4.3 Size Standards

DNA size standards

For agarose gels, Hyper Ladder I® (Bioline) was used (m.wt. 200, 400, 600, 800 1000, 2000, 2500, 3000, 4000, 5000, 6000, 8000, 10000 bp)

Protein size standards

High molecular weight range (SIG) (m.w. 36, 45, 55, 66, 84, 97, 116 and 205 kDa)

Low molecular weight range (SIG) (m.w. 20, 24, 29, 36, 45 and 66 kDa).

Lyophilised standards (SIG) were reconstituted with 100 µl of sterile 18.2 MΩ/cm H_2O to give a final concentration of ~2.0-3.5 mg / ml and separated into 4 µl aliquots and stored at –20°C.

2.1.4.4 $MgCl_2$-$CaCl_2$ solutions

For the preparation of chemically competent *E. coli* cells for transformation

$MgCl_2$ (SIG) (80 mM) solution

(3.3 g/ 200 ml 18.2 MΩ/cm H_2O)

$CaCl_2$ (SIG) (20 mM) solution

(2.94 g/ 200 ml 18.2 MΩ/cm H_2O)

The two solutions were mixed in a 1:1 ratio and separated into 20 ml aliquots followed by autoclave sterilization. The aliquots were then stored at 4°C.

For final re-suspension of chemically competent cells 1 M $CaCl_2$ was prepared by dissolving 14.7 g of $CaCl_2$ in 100 ml 18.2 MΩ/cm H_2O. The solution was separated in 1 ml aliquots and stored at 4°C.

2.1.5 SDS-PAGE reagents and buffers

12% (w/v) acrylamide resolving gel components

Component	Volume
40% (w/v) solution (37.5:1 acrylamide: bisacrylamide) (SIG)	3 ml
Buffer B*	2.5 ml
18.2 MΩ/cm H_2O	4 ml
Ammonium persulphate (APS) 10% (w/v) (SIG)	50 µl
N,N,N',N'-Tetramethylethylenediamine (TEMED) (SIG)	10 µl

*Buffer B

Component	Volume (ml)/ 100 ml 18.2 MΩ/cm H_2O
TRIS-HCl (2 M) pH 8.8 (SIG)	75
10% (w/v) SDS (SIG)	4

Acrylamide stacking gel components

Component	Volume
40% (v/v) solution (37.5:1 acrylamide: bisacrylamide)	0.5 ml
Solution C*	1 ml
18.2 MΩ/cm H_2O	2.5 ml
10% (w/v) APS (SIG)	30 µl
TEMED (SIG)	10 µl

5 µl of 2% (w/v) Bromophenol blue dye was added to the stacking gel to give a blue colour to distinguish between wells while loading samples.

*Buffer C

Component	Volume (ml)/ 100 ml 18.2 MΩ/cm H_2O
TRIS-HCl (2 M) pH 8.8 (SIG)	50
10% (w/v) SDS (SIG)	4

SDS - PAGE running buffer 10x (Stock)

Component	Amount (g)/L 18.2 MΩ/cm H_2O
TRIS	0.3
Glycine (SIG)	144
SDS (SIG)	10

The pH of the TRIS-HCl and glycine was adjusted to pH 8.0 in a volume of ~900 ml prior to the addition of SDS.

Running buffer was made up at 10 x stock concentration and diluted 1:10 with 18.2 MΩ/cm H_2O prior to use.

SDS - PAGE sample loading dye

Component	Volume (ml)/10 ml
TRIS-HCl (60 mM) pH 6.8	0.6
50% (v/v) Glycerol (SIG)	5.0
10% (w/v)SDS	2.0
β-Mercaptoethanol (14.4 mM) (SIG)	0.5
Bromophenol blue (0.1%) (w/v) (SIG)	1.0
18.2 MΩ/cm H_2O	0.9

Loading dye was stored at –20°C in 0.4 ml aliquots.

Solublising SDS - PAGE sample buffer (cracking buffer)

2.4 g Urea (SIG) was added to 7.6 ml SDS-PAGE sample loading dye (section **2.1.3.4**) and stored at 4°C.

Coomassie blue SDS-PAGE gel staining solution

Component	Per L 18.2 MΩ/cm H_2O
Coomassie Blue R-250 (FIS)	1 g
Glacial acetic acid	100 ml
Methanol (FIS)	450 ml

Coomassie blue SDS-PAGE gel destaining solution

Component	volume (ml)/ L 18.2 MΩ/cm H_2O
Acetic acid	100
Methanol	100

2.1.6 Buffers for protein purification

2.1.6.1 Ni^{2+} affinity chromatography purification buffers (for automated purification, gradient elution)

Start buffer (Na_2HPO_4 (50 mM), NaCl (0.5 M), imidazole (10 mM)) pH 7.4.

Component	Amount (g)/ L 18.2 MΩ/cm H_2O
Na_2HPO_4 (SIG)	7.1
NaCl	29.2
Imidazole (ACR)	0.68

Elution buffer (Na_2HPO_4 (50 mM), NaCl (0.5 M), imidazole (500 mM)), pH 7.4

Refer recipe for start buffer except 0.68 g Imidazole was replaced with 34 g Imidazole.

2.1.6.2 Buffers for MS analysis of digested peptides

LC-MS grade H_2O and Acetonitrile (ACN) (FIS) were used to prepare buffers for mass spectrometry analysis of peptides. Buffers were made up fresh prior to use.

Start buffer A

Component	% (v/v)
18.2 MΩ/cm H_2O	95
ACN	5
Formic acid	0.1

Start buffer B

Component	% (v/v)
18.2 MΩ/cm H_2O	5
ACN	95
Formic acid	0.1

2.1.7 In-gel protein digestion reagents

100 mM NH_4HCO_3 (pH 8.0)

79 mg NH_4HCO_3 was dissolved in 10 ml 18.2 MΩ/cm H_2O. A 1:2 dilution of 100 mM NH_4HCO_3 was prepared in 18.2 MΩ/cm H_2O to prepare 50 mM NH_4HCO_3.

Trypsin solution

A 1 µg/µl stock solution of Trypsin (Trypsin Gold, Mass spectrometry grade, Promega) was prepared by dissolving 100 µg of the lyophilised powder in 100 µl 50 mM glacial acetic acid. Stock solutions were either stored at -20°C for use within one month or at -80°C for future use.

2.1.8 Yeast and *E. coli* (DE3) strains

All yeast strains were sourced from suppliers as detailed in Table **2.3.**

Organism	Type	Supplier and collection no.	GenBank Accession Nos. of UGTs
Saccharomyces cerevisiae (Sc) strain S288C	Yeast	ATCC 204508	NP_013290
Pichia pastoris (Pp) strain GS115	Yeast	ATCC 201178	AAD29570
Pichia guilliermondii (Pg)	Yeast	DSM 6381	XP_001482815
Kluyveromyces polysporus (Kp)	Yeast	DSM 70294	XP_001646967
Pichia angusta (Pa)	Yeast	DSM 70277	ABO31066
Kluyveromyces lactis (Kl)	Yeast	DSM 70799	XP_452287
Debaromyces hansenii (Dh)	Yeast	CBS 767[a]	XP_460322

Table 2.3 Summary of yeast host strains used in this study. [a] Received as a gift from Agriculture Research Service Culture Collection, U.S.

Stocks of all *E. coli* (DE3) strains were maintained at –80°C in 50% (v/v) glycerol and were used as detailed in **Table 2.4.** Antibiotic requirements for each strain used are listed in **Table 2.5.**

***E. coli* strain and source**	**Application**	**Genotype [b]**
TOP10 (INV)	Cloning host	F⁻ *mcr*A Δ(*mrr*⁻ *hsd*RMS⁻*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*araleu*) 7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G
DE3 strains *:*		
BL21 (NOV)	Optimising soluble protein expression	F⁻ompT hsdS_B (r_B⁻m_B) gal dcm (DE3)
B834 (NOV)		F⁻ *ompT hsd*SB (*r*B⁻ *M*b⁻) *gal dcm met* (DE3)
C41 (NOV)		F⁻ *ompT hsdS*$_B$ (r_B⁻ m_B⁻) *gal dcm* (DE3)
C43 (NOV)		F-*ompT gal hsdSB* (rB-mB⁻) *dcm lon* λ DE3
BL21-Codon Plus –RP (CP-RP) [c]		B F⁻ *ompT hsdS*(rB⁻ mB⁻) *dcm*⁺ Tetr *gal* λ(DE3) *endA* Hte [*argU proL*Camr]
BL21-Codon Plus –RIPL (CP-RIPL)[c]		B F⁻ *ompT hsdS*(rB⁻ mB⁻) *dcm*⁺ Tetr *gal* λ(DE3) *endA* Hte [*argU proL* Camr] [*argU ileY leuW* Strep/Specr]

Table 2.4 *E. coli* strains and their applications in this study. ([b] pET system manual, 1999). [c] Glycerol stocks provided by Structural Biology Lab, York University

***E. coli* strain**	**Antibiotic Required for growth (liquid/solid media)**
TOP 10	None
BL21	None
B834	None
C41	None
C43	None
CP-RP	chloramphenicol
CP-RIPL	Chloramphenicol in liquid media prior to induction with IPTG and for preparation of chemically competent cells Chloramphenicol and streptomycin for growth in solid and liquid media (preparation of glycerol stocks etc.)

Table 2.5 Antibiotic growth requirements for *E. coli* (DE3) strains.

2.1.9 Enzymes

All enzymes, buffers and reaction co-constituents were maintained at –20°C. Enzymes and reaction constituents are listed in **Table 2.6.**

Enzyme	Suppliers	Co-constituents
Polymerases:		
T4 DNA Polymerase	NOV	Reaction buffer (recipe undisclosed by manufacturer)
KOD Hot Start DNA Polymerase (kit)	BL	25 mM $MgSO_4$, reaction buffer (recipe undisclosed by manufacturer)
Taq DNA Polymerase (kit)	STR	50 mM $MgCl_2$, NH_4 reaction buffer (160 mM $(NH_4)_2 SO_4$, 670 mM TRIS-HCl (pH 8.8)
Mutazyme II® DNA polymerase, EZClone DNA polymerase (kit)		Supplied reaction buffers (recipes undisclosed by manufacturer)
Restriction enzymes *Xba*I and *Xho*I	NEB	Bovine serum albumin (BSA)[e], reaction buffer 2 (10 mM TRIS-HCl, 50 mM NaCl, 10 mM $MgCl_2$, 1mM Dithiothreitol (DTT), pH 7.9 at 25°C).
Lysozyme (from hen egg white)	FIS	None
Lyticase (*Arthrobacter luteus*)	SIG	Sorbitol buffer (1M sorbitol (FIS); 100 mM sodium EDTA; 14 mM β-mercaptoethanol)
RNase	SIG	None
*Dpn*I	NEB	None
Trypsin (Mass spectrometry grade)	Promega	50 mM glacial acetic acid.

Table 2.6 Enzymes and co-constituents. [e]BSA (NEB) was supplied at 10 mg/ml (100x) and was diluted to 10x with 18.2 MΩ/cm H_2O prior to use.

2.1.10 Kits

The molecular biology kits employed in this study are listed in Table **2.7.** All kits and components were stored at room temperature, except the cell re-suspension buffer of plasmid purification kits, which was maintained at 4°C.

Kit	Application
DNeasy Blood and Tissue Kit® (Qiagen)	Extraction of genomic DNA from host
Miniprep kit (NZytech)	Extraction of plasmid DNA (pDNA) from *E. coli* (DE3)
PCR purification kit (Qiagen)	Purification of double stranded DNA (dsDNA) from a PCR.
Nucleotide removal kit (Qiagen)	Removal of nucleotides following LIC-T4-polymerase reactions.

Table 2.7 Molecular biology kits and applications.

2.1.11 Oligonucleotide primers (MWG)

Oligonucleotides were received as lyophilised pellets, which were centrifuged at 13000 x g for 4 min and were made up to a stock concentration of 100 pmol/µl (100 µM) with the appropriate volume of 18.2 MΩ/cm H_2O as indicated by the supplier. Forward (CGCGCCTTCTCCTCACTGTTCCA) and reverse (TGTGGTGGTGGTGATGATGGCTG) primer sequences were used for PCR

amplification of the LIC vector. Primer stock concentration was 100 μM (each of forward and reverse). All primers were maintained at -80°C.

2.1.12 Vectors

pET-YSBLIC (LIC) vector was used for cloning all gene fragments, pET- 28a was used as a control for transformation and in high throughput screening procedures.

2.1.13 Substrates

All acceptor and donor substrates are listed in **Table 2.8 (a) and (b)** respectively. Acceptor stock solutions of 2-4 mM were prepared in 100% (v/v) ethanol. Donor sugar solutions were prepared as 40 mM stocks in sterile 18.2 MΩ/cm H_2O and were separated as 50 μl aliquots and stored at -20°C.

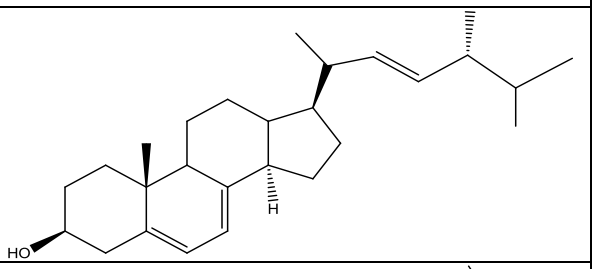
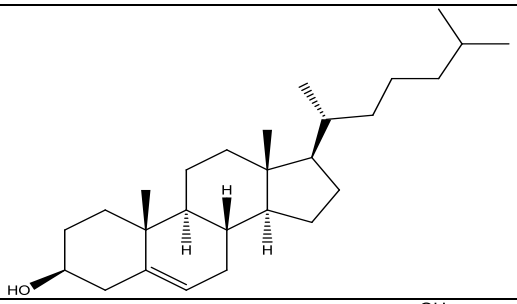
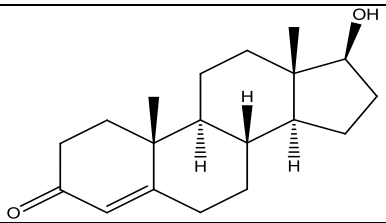
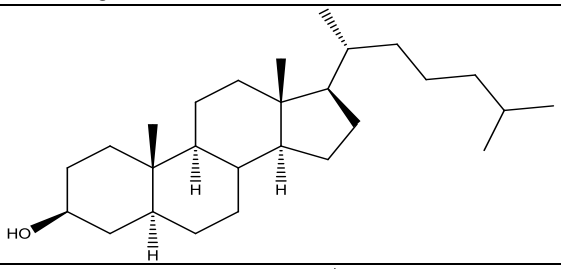
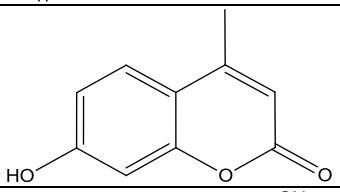
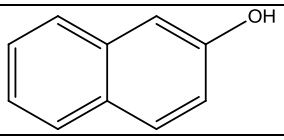
Acceptor substrate	Structure
ergosterol (SIG)	
cholesterol (SIG)	
testosterone (SIG)	
5-alpha-cholestan-3-beta-ol (dihydrocholesterol) (AE)	
4-methylumbelliferone (4-MU) (SIG)	
β-naphthol (SIG)	

Table 2.8 (a) Chemical structures of acceptor aglycones.

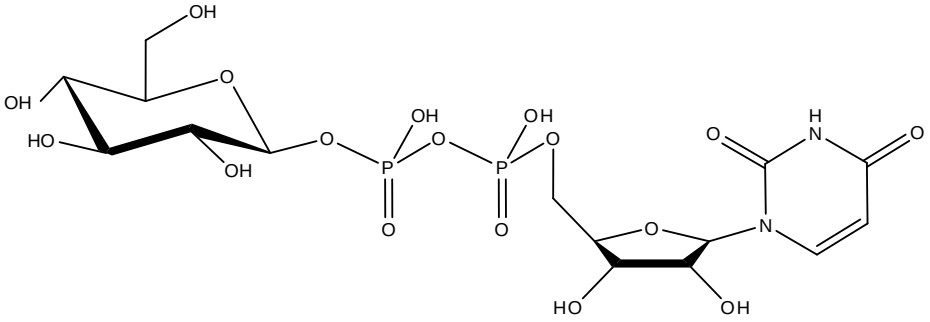
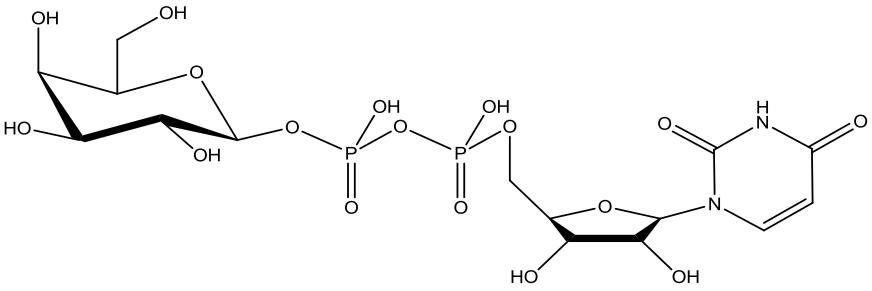
Donor substrate	structure
UDP-glucose (*S. cerevisiae*) (UDPG) (SIG)	
UDP-galactose (UDPGA) (SIG)	

Table 2.8 (b) Chemical structures of donor sugars.

2.2 Methods

All basic molecular biology techniques including cloning and protein expression and purification were followed as described in Sambrook and Russell (Editions 1, 2 and 3; 2001).

2.2.1 DNA Methods

2.2.1.1 Growth of yeast starter cultures for DNA extraction

Growth conditions of each of the yeast strains are detailed in **Table 2.9**. A loopful of each of the yeast cultures provided was inoculated in 5 ml appropriate media in a sterile glass universal and left to grow overnight at 100 rpm at the corresponding temperature. Recipes of media are detailed in section **2.1.2.1**.

2.2.1.2 Growth of *E. coli* (DE3) starter cultures for plasmid DNA (pDNA) extraction

E. coli (DE3) were grown in 28 ml sterile glass universals containing 5/10 ml of LB broth (depending upon the volume of inoculum required for the next stage) supplemented with the appropriate antibiotic, the resistance gene for which is encoded on the plasmid to be purified or expressed. For crude extraction of pDNA, cultures were inoculated from a single colony on an agar plate using a sterile wire loop whereas for standard extraction of pDNA (spin DNA), cultures were inoculated from an agar plate, glycerol stock or liquid culture using a sterile wire loop. Cultures were grown overnight at 37°C with orbital shaking at 200 rpm.

2.2.1.3 Growth of *E. coli* (DE3) for the preparation of cell free extracts (CFEs)

A 5/10 ml starter culture (as described above) for clones harbouring recombinant plasmids was prepared. Starter cultures were then inoculated at a 1% (v/v) ratio to a conical flask containing LB broth supplemented with the appropriate antibiotic (See **Table 2.5** for antibiotic growth requirements of *E. coli* strains). Small scale expression was performed to study protein expression level profiles when using different DE3 hosts or auto-induction media (**Appendix C.**) and scaled up to achieve large amounts of protein for further analysis. **Table 2.10** lists the scales of expression used and the volume of broth.

Organism	Media	Growth temperature
Sc	YEPD	30°C
Pp	YEPD	30°C
Pg	YM	26°C
Kp	393/YPD	25°C
Pa	186	25°C
Kl	186	25°C
Dh	YM	25°C

Table 2.9 Growth conditions for yeast strains for genomic DNA extraction.

Scale of expression	Volume of broth (ml)	Maximum capacity of conical flask (ml)	Volume of inoculum (ml)
Small	50	250	0.5
Large	1000	2500	10

Table 2.10 Scales of protein expression.

The inoculated culture was grown and allowed to grow at 37°C orbital shaking at 200 rpm for approximately 2-3 h till exponential phase with an $O.D_{600nm}$= 0.7 read by spectrophotometer previously zeroed with LB media. The cells were then induced for protein expression by the addition of 240 µg/ml IPTG (final concentration) and allowed to grow overnight at 23°C shaking at 100 rpm.

2.2.1.4 Extraction and purification of genomic DNA from yeast cultures

Genomic DNA was extracted from overnight grown yeast cultures (section **2.2.1.1**) using a kit (**Table 2.7**). The recipes for the buffers used in the kit are undisclosed by the manufacturer.

Overnight cultures (0.5 ml) were harvested in a 1.5 ml micro-centrifuge tube by centrifuging for 10 min at 5000 x *g*. The resulting supernatant was discarded and the pellet re-suspended in Sorbitol buffer (Recipe in **Table 2.6**) followed by the addition of 200 U lyticase (430 U/ml) and incubation at 30°C for 30 min. The cells were centrifuged for 10 min at 300 x *g* and re-suspended in 180 µl buffer ATL (cell re-suspension buffer). 20 µl proteinase K was then added to the cell mixture mixed by vortexing prior to incubation at 55°C for approx. 3 hr to ensure

complete tissue lysis. Cells were vortexed for 15 s followed by the addition of 200 µl buffer AL (for precipitation of chromosomal DNA), mixed by vortexing and incubated at 70°C for 10 min. 200 µl ethanol (100% v/v) was added to the sample and mixed thoroughly by vortexing. The resulting mixture from step was then transferred to a DNeasy Mini spin column placed in the provided collection tube. Cells were then centrifuged for ≥6000 x *g* for 1 min and the flow-through discarded. Sample was then washed by the addition of 500 µl buffer AW1 followed by an additional wash with 500 µl buffer AW2 to the spin column, both washes were performed by centrifugation for 3 min at 20,000 x *g* and the flow-through discarded in both cases. The DNeasy column was placed in a clean microcentrifuge tube and 100 µl buffer AE (5 mM TRIS-HCl, pH 8.0) was then added directly in the centre of the DNeasy membrane. This was followed by an incubation step at room temperature for 1 min and centrifugation at 14, 000 x *g*. This step was repeated and the resulting 200 µl of eluant was stored at -20°C until further use.

2.2.1.5 Extraction of pDNA from bacteria

Crude extraction of pDNA from bacteria using STET Buffer (to screen for plasmids carrying inserts)

All centrifugation steps were performed at 14,000 x *g*. A 3 ml *E. coli* culture (section **2.2.1.2**) was pelleted by repeated centrifugation for 1 min. The supernatant was discarded followed by pulsed to remove any residual supernatant. The resulting pellet was re-suspended in 150 µl STET Buffer (section **2.1.4.2**) ensuring complete re-suspension of cells by vortexing. 10 µl of 10 mg/ml lysozyme (prepared fresh on the day) was then added to cells and samples were

mixed by gentle inversion of the tube about 6-7 times. The cells were incubated at room temperature for 15 min and immersed in a boiling water bath for 40 s. Samples were immediately centrifuged at 14000 x g for 15 min. The resulting viscous pellet was then removed gently using a pipette tip and discarded. 150 µl chilled isopropanol was then added to the remaining supernatant and mixed. The cells were incubated at -20°C for 1 h to allow precipitation of DNA followed by centrifugation of cells for 5 min. The resulting supernatant was discarded and the pellet pulsed to ensure removal of any residual supernatant. The pellet was washed with 500 µl 100% (v/v) ethanol and centrifuged for 3 min. The resulting supernatant was discarded and the residual supernatant was carefully removed using the pipette. Samples were then air-dried for 10 min to allow evaporation of any latent ethanol. The dried pellet was re-suspended in 30 µl of TE Buffer (section **2.1.4.1**) containing 100 µg/ml RNase.

Standard extraction of pDNA from bacteria (Spin DNA)

Spin DNA was extracted from bacterial cultures (section **2.2.1.2**) using a kit (**Table 2.7**). The recipes for buffers used are undisclosed by the manufacturer. All centrifugation steps were performed at 11,000 x *g* unless stated otherwise. A 3 ml (from 5 ml) volume of overnight grown bacterial cell culture was pelleted in a 1.5 ml micro-centrifuge tube by repeated centrifugation for 30 s and the supernatant discarded. The pellet was then pulsed and residual supernatant removed. The pellet was then re-suspended by vortexing in 250 µl of buffer A1 (cell re-suspension buffer). 250 µl of buffer A2 (cell lysis buffer) was then added and the sample was mixed by gentle inversion 5-6 times, prior to incubation at room temperature for 5 min. 300 µl of buffer A3 (precipitation of proteins and chromosomal DNA) was then added to the sample and the sample was again

mixed by gentle inversion 5-6 times Precipitated material was then separated by centrifugation for 10 min and the supernatant was transferred to a spin column placed in a 2 ml receptacle collecting tube. The DNA was then loaded onto the filter by centrifugation for 1 min and the flowthrough discarded and washed with 500 µl buffer AY, followed by an additional wash with 600 µl of buffer A4 (50% ethanol), both washes were performed by centrifugation at 14000 x *g* for 1 min and in each case the flowthrough discarded. Prior to elution, the empty spin column was centrifuged for 2 min in order to remove any residual ethanol. The spin column was then transferred into a sterile 1.5ml micro-centrifuge tube and the DNA eluted from the filter by centrifugation at 14000 x *g* for 1 min after the addition of 50 µl of AE (5 mM TRIS-HCl, pH 8.0) and incubation at room temperature for 1 min. For extraction of DNA for sequencing reactions, the same procedure was followed except: 9 ml overnight starter culture was pelleted (instead of 3 ml) and twice the volume of each buffer was used in each step. In the final step, DNA was eluted in 50 µl sterile 18.2 MΩ/cm H_2O (pH 8.0).

2.2.1.6 PCR Purification (PCR products, *Dpn*I digested LIC vector)

Purification of double stranded DNA (dsDNA) following a PCR was achieved using a kit (**Table 2.7**). The method is detailed as follows: All centrifugation steps were performed at 13,000 x *g*. Recipes for buffers used are not known.

5 volumes of buffer PB were added to 1 volume of the PCR sample and mixed. A spin column was placed in a 2 ml collection tube following the application of the sample to the column. The column was then centrifuged for 60 s to bind the DNA. The flow-through was discarded and the column was placed back into the same tube. 0.75 ml buffer PE (wash buffer) was added to the spin column and

centrifuged for 60 s. The flow-through was discarded and the column placed backed into the same tube. The column was centrifuged for an additional 1 min to evaporate the residual ethanol from buffer PE. The spin column was placed in a sterile 1.5 ml microcentrifuge tube. To elute DNA 50 µl sterile 18.2 MΩ/cm H_2O (pH 8.0) was added to the centre of the column membrane and the column was centrifuged again for 1 min. For elution of LIC vector following PCR, the elution step was performed using 100 µl TRIS-HCl (pH 8.0) and was repeated.

2.2.1.7 AGE and gel visualization

Agarose (1% w/v) gels were prepared in TAE buffer (section **2.1.3.1**) in a total volume of 30/120 ml (depending upon the size of the tank used). The mixture was then boiled over a Bunsen burner until the agarose powder had completely dissolved in the buffer. This was confirmed by holding the conical flask against light to check no agarose particles remained. Once the mixture cooled to around 60°C, it was poured into an agarose gel casting tray, a well comb inserted and the gel allowed to set for approx. 20 min. The casting tray with the 'set' agarose gel was then transferred into a gel electrophoresis tank and the tank filled with TAE buffer till it submerged the gel. The samples were then loaded onto the gel along with a DNA size standard (section **2.1.4.3**). The electrophoresis tank was then connected to a power pack and electrophoresis carried out at 120 mA and 200 V for 30 min when using the small tank and 50 min while using the large tank. The gel was then transferred to a staining tray and filled with freshly made 10 µg/ml EtBr solution (section **2.1.3.2**) until the solution just submerged the gel. The tray was placed on a bench-top orbital shaker and the gel allowed staining for 15 min at 100 rpm. After staining, the EtBr solution was disposed by charcoal filtration and the gel rinsed with distilled H_2O. After electrophoresis, the gel was visualised

using a gel documentation system using Quantity One™ software. Hard copies of the gel picture were produced using a Mitsubishi Video Copy Processor (Model P91 attached to the gel doc system), with Mitsubishi thermal paper (K65HM-$_{CE}$ / High density type, 110 mm x 21 m).

2.2.1.8 Preparation of competent *E. coli* (DE3) cells for transformation

Preparation of chemically competent *E. coli* (DE3)

From a glycerol stock of *E. coli* (DE3) cells, a scrapping of cells was taken using a sterile loop and streaked onto a 4% w/v agar plate supplemented with the appropriate antibiotic if required (**Table 2.5**) and incubated overnight at 37°C. Approximately 1 cm of cell scrapping was then taken with a sterile loop from the agar plate and inoculated into 50 ml LB broth (section **2.1.2.1**). For *E. coli* (DE3) strains Codon Plus CP (RP) and CP (RIPL), appropriate antibiotic was added to the liquid media prior to inoculation (**Table 2.5** lists antibiotic growth requirements of different strains). The cultures were incubated at 37°C orbital shaking at 200 rpm for approx. 2 h till O.D$_{600nm}$ was between 0.35 and 0.40 as read by a spectrophotometer previously blanked with LB broth. At this point, the cultures were kept on ice for approx. 30 min and pelleted by centrifugation in two pre-chilled 25 ml sterile polystyrene universals at 2000 x *g* for 10 min at 4°C. The supernatant was discarded and the tubes were briefly inverted on tissue to drain the liquid ensuring that cells were not left off ice for more than a few seconds. One of the pellets was re-suspended by flicking the tube gently in 15 ml of ice cold sterile pre-chilled $MgCl_2$-$CaCl_2$ solution (section **2.1.4.4**). The re-suspended pellet was then transferred to the second tube and the pellet re-suspended similarly. The cells were centrifuged again and the pellet drained as described

earlier. The pellet was then re-suspended in 1 ml ice cold 100 mM $CaCl_2$ (section **2.1.4.4**). Chemically competent cells for expression hosts (BL21 etc) could be prepared and frozen at -80°C in 15% (v/v) glycerol for future use.

Preparation of electrocompetent *E. coli* (DE3) cells for transformation

An agar plate of the glycerol stock of the required cell type was prepared and incubated overnight at 37^{o}C (see preparation of chemically competent *E. coli*). After overnight growth, using a sterile loop, approximately 1 cm of cell scrapping was inoculated into 50 ml of sterile pre-warmed low salt LB broth (section **2.1.2.1**) in a 250 ml conical flask. The culture was then grown at 37 ^{o}C with orbital shaking at 200 rpm until the cells reached an $O.D_{600nm}$ of 0.5~0.7. The culture was then placed on ice for 30 minutes and centrifuged at 4000 x g for 15 min at 4^{o}C and the supernatant discarded. The pellet was then gently re-suspended (by gentle vortex, never more than 3 sec at a time off ice) in 25 ml of ice cold filter sterilized 18.2 MΩ/cm H_2O. The suspended cells were then centrifuged as before and the supernatant discarded. The cell pellet was then re-suspended (by gentle vortex never more than 3 sec at a time off ice) in 12 ml of ice cold filter sterilized 18.2 MΩ/cm H_2O. The centrifugation and re-suspension was repeated twice but with the re-suspension volume decreasing each time with the next volume being 15 ml followed by re-suspension in 2 ml. A final centrifugation and re-suspension step followed with a volume of 0.2 ml of ice cold filter sterilized 18.2 MΩ/cm H_2O after which the cells were separated into pre-chilled microcentrifuge tubes in 40 µl aliquots.

2.2.1.9 Spectrophotometric quantification of DNA

Plasmid DNA was quantitatively determined by transferring 60 µl neat sample to a sterile disposable plastic cuvette (UVette®, Eppendorf) fitted in an appropriate adapter for the UV-vis spectrophotometer (Ultrospec 2000 spectrophotometer, Pharmacia Biotech). In cases where the sample was highly concentrated or < 60 µl in volume, appropriate dilutions were prepared in 18.2 MΩ/cm H_2O. A_{260} value was then recorded by a UV-visible spectrophotometer previously zeroed with water. The optimal wavelength for contaminating protein was assumed to be 280 nm. The concentration and purity of the DNA was then determined as detailed below:

[dsDNA] µg/µl = (A_{260}) x 50 x (dilution factor) / 1000.

DNA purity ratio = (A_{260})/ (A_{280})

A DNA purity ratio of >1.8 was accepted as minimal protein contamination.

2.2.1.11 Restriction endonuclease digests

Restriction enzyme digestion of pDNA (section **2.2.1.5**) was performed based on the knowledge of restriction sites present at the 5' and 3' ends of the gene of interest as described in section **2.2.1.11**. The following were added to a 1.5 ml microcentrifuge tube (**Table 2.11**) for restriction endonuclease reactions. Restriction enzymes were added in the end and mixed by gently swirling the pipette tip.

Component	**Volume** (µl)	**Final concentration**
pDNA (30-120 ng/µl)	5	15-60 ng/µl
*Xba*I (20,000 U/ml)	0.5	10 U
*Xho*I (20,000 U/ml)	0.5	10 U
Restriction Enzyme Buffer 2 [1]	1	
10 x BSA [1]	1	1x
18.2 MΩ/cm H_2O to a final volume of 10 µl	2	

Table 2.11 Restriction enzyme digestion conditions of plasmid DNA. [1] Recipes in **Table 2.6.**

Each reaction was incubated for 2 h in a 37^{o}C water bath. 2 µl of DNA loading dye was added to the 10 µl of each digest and analysed by gel electrophoresis as on a 1% (w/v) agarose gel.

2.2.1.12 Transformation of recombinant plasmids into *E. coli* (DE3)

Transformation using chemically competent *E. coli* (DE3)

Transformations of insert-LIC constructs and r-plasmid DNA were performed using prepared chemically competent *E. coli* cells (see section **2.2.1.8** for preparation of competent *E. coli*). 5 µl pDNA was added to 50 µl chemically competent *E. coli* cells (if made fresh on the day) or 75 µl cells (previously made and frozen in 15% (v/v) glycerol).) in a 1.5 ml microcentrifuge tube and mixed gently with a pipette tip. The mixture was then incubated on ice for 30 min. The cells were then transferred to a water bath at 42°C and heat shocked at this

temperature for 90 s and immediately transferred on ice for 2 min. The cells were then recovered by the aseptic addition of 200 µl of sterile LB media (section **2.1.2.1**), and incubated at 37°C for 1 h. Recovered cells were then aseptically plated on agar plates supplemented with kanamycin (see **Table 2.2**). Plates were then incubated for 16 h at 37°C to allow the growth of bacterial colonies.

Apart from the conventional expression strain, BL21 (DE3), other host *E. coli* strains were also used with antibiotic selection requirements as stated in **Table 2.3.**

Transformation of mutant library DNA into electrocompetent *E. coli* cells (by electroporation)

Transformations were performed using prepared electrocompetent cells. 40 µl aliquots of cells were placed on ice and 2 µl of ligation or pDNA was added and mixed gently with a pipette tip. The cells and DNA mix was incubated on ice for at least 1 min. The mix was then transferred to an ice cold sterile 2 mm electroporation cuvette (BR) and the contents tapped gently down to the bottom of the cuvette until it entirely covered the bottom of the cuvette. The outside of the cuvette was dried thoroughly and the cuvette placed into a BioRad ShockPad and the lid closed. The BioRad PulseXcell was set to 2.5 kV and the voltage applied across the cuvette. Immediately following electroporation 1 ml of LB broth was added to the cuvette and transferred using a sterile glass pipette to a sterile 1.5 ml microcentrifuge tube. The cells were incubated at 37 ^{o}C for at least 1 hour and then plated onto an appropriate antibiotic LB agar plate. Plates were then incubated inverted for ~16 h at 37 ^{o}C.

2.2.1.13 Plating bacteria

Prior to plating, LB agar plates were surface dried by placing open face down at 65°C for 10 min. A glass spreader was sterilised by immersion in 70% (v/v) ethanol, followed by removal of excess ethanol by passing quickly through a blue Bunsen flame, and used to spread 50-200 µl of bacterial suspension evenly over the surface of the agar. When all the liquid had been absorbed into the agar, the plate was incubated inverted at 37°C for ~ 16 h.

2.2.2 Protein Methods

2.2.2.1 SDS-PAGE electrophoresis

SDS – PAGE was conducted according to the method of Laemmli (Laemmli 1970).

SDS - PAGE gels were cast between two clean glass plates of dimensions (10.1 x 7.2 and 10.1 x 8.2 cm) separated by a 1 mm spacer ridge on the larger of the two plates. The plates were clamped together and checked to ensure the bottoms were flush, prior to securing the gel with vertical downward pressure in a casting stand on a rubber gasket. The components for the resolving gel (section **2.1.5**) were then added together, mixed, and then pipetted (using a 5 ml Gilson) into the space between the plates, until the solution was approximately 2 cm below the top of the smallest plate. In order to prevent oxygen from inhibiting polymerisation, a small volume of 18.2 MΩ/cm H_2O was then added to the top of the solution. The resolving gel was then allowed to polymerise for 20 min. Once set, the H_2O on top of the gel was removed by adsorbing with filter paper, and the components of

the stacking gel were added (section **2.1.5**), mixed, and pipetted up to the top of small plate. A 10 toothed comb was then immediately inserted into the gap between the plates displacing some of the un-polymerised stacking gel. The stacking gel was then allowed to polymerise for 20 min prior to removal of the comb to reveal the wells, which were subsequently rinsed with H_2O to remove any gel debris. The gel was then placed vertically into the buffer tank secured in a holder with, another gel constructed in exactly the same manner secured to the opposite side. SDS-PAGE 1 x running buffer (section **2.1.5**) was then poured into the tank ensuring the level of buffer in the space between the two gels was higher than the level on the outside of the gels. Samples were then prepared by the addition of 5 µl of 6 x loading buffer (section **2.1.5**) to 20 µl of sample, which was then pulsed briefly prior to boiling for 2 min. Size standards were prepared in the same way, except 8 µl of loading dye was added to 4 µl of size standard (section **2.1.4.3**). All samples and standards were then centrifuged at 14000 x *g* for 1 min prior to loading 20 µl of all samples and 10 µl of all standards into the wells using a Hamilton syringe. Finally the pair of gels was run at 120 mA, 200 v for 45 min. Gels were carefully removed and transferred to a staining tray after which Coomassie blue staining solution (section **2.1.5**) was poured till it submerged the gels. The tray was placed on a bench-top orbital shaker and the gel allowed staining for 15 min at 100 rpm after which the staining solution was removed and disposed by funnel filtration and the gels rinsed with distilled H_2O. The tray with gels was filled with gel de-staining solution (section **2.1.5**) and left shaking overnight at 100 rpm to completely remove the staining solution. Stained gels were then analysed by a densitometer (section **2.1.1.7**).

2.2.2.2 Preparation CFEs

500 µl of each of the induced overnight *E. coli* cultures (section **2.2.1.3**) was centrifuged at 4000 x *g* for 1 min, the supernatant decanted. The pellet was maintained as the insoluble protein fraction and was resolubilised in 300 µl of solubilising buffer (section **2.1.5**).

The overnight cultures were then concentrated at 4000 x *g*, 4^{o}C for 15 min, the supernatant decanted and the pellets re-suspended in 5 ml start (section **2.1.6**). Cells were then lysed in 5 ml volumes by sonication on ice at amplitude 14000 microns for 2 min in 10 s bursts with 10 s intervals. The lysate was then decanted into centrifuge tubes and centrifuged at 4°C for 20 min at 24000 x *g*. After centrifugation the soluble fraction was maintained on ice as the CFE. The CFE and insoluble protein fractions were then analysed for protein expression by SDS-PAGE (section **2.2.2.1**).

2.2.2.3 Protein Purification via Nickel affinity chromatography

N-terminally his-tagged protein was purified from cell free extract (section **2.2.2.2**) by automated Immobilised Metal Affinity Chromatography (IMAC) using automated gradient elution using a Ni Sepharose High Performance resin (AB). This was performed using the Akta Prime Plus 5.0 FPLC system™. In this programmed method, the gradient elution column was prepared by calibration with 100 ml start buffer (section **2.1.6**). 5-10 ml of CFE was then loaded onto the column. The protein was finally eluted at a flow rate of 2 ml/min by passing through 200 ml elution buffer (section **2.1.6**) with a linear imidazole gradient (10-500 mM). The elution time of the protein was monitored at 280 nm using an inline UV spectrometer connected to the FPLC after which fractions (5 ml) were

collected by a fraction collector. To confirm purity, a 20 µl aliquot from each 5 ml fraction thought to contain the target protein was analysed by SDS-PAGE (section **2.2.2.1**) and the pure fractions pooled.

2.2.2.4 Concentration and buffer exchange of protein

Concentration of protein was performed at 4°C by centrifugation at 4000 x *g* in 30 kDa cut-off concentrator units. Solution was washed three times with start buffer (section **2.1.6**) and stored at 4°C in 5 mM TRIS-HCl (pH 8.0) at a concentration of 1 mg/ml for future use.

2.2.2.5 Spectrophotometric quantification of protein

A 60 µl protein sample aliquot was transferred to a sterile synthetic Quartz cuvette (Uvette, Eppendorf) and absorbance at 280 nm was recorded by a UV-visible spectrophotometer after the instrument was zeroed with 5 mM TRIS-HCl (pH 8.0). The concentration of proteins was determined as detailed below:

[Protein] µg/µl = (A_{280}) x m.wt of protein x (dilution factor)/Molar extinction coefficient

2.2.3 UGT activity assays using UDP-[U-^{14}C] glucose (UDP*G*) followed by liquid scintillation counting (LSC) of extracted glycosides.

All pre-assay enzyme dilutions were prepared in 100 mM TRIS-HCl; pH 8.0. A 2 mM stock solution for testosterone and 1 mM stock solution for cholesterol was prepared in ethanol (99% v/v). UDP*G* (PE) was supplied at 0.02 mCi/ml (specific activity 302 mCi/mmol) in 7:3 (v/v) ethanol: water. A 1:5 dilution of the stock

was prepared in H_2O for each ~ 5 µl dilution = 50,000 dpm (0.07 µM). This was confirmed by liquid scintillation counting (LSC). Diluted solution was separated as 200 µl aliquots and stored at -20°C.

All reactions were set in triplicate along with a blank which consisted of 75 µl buffer and a negative control consisting of all constituents except the acceptor substrate which was substituted with 18.2 MΩ/cm H_2O. Assay constituents using purified GTs and CFEs are listed in **Table 2.12 (a)** and **(b)** respectively. All components except UDP*G* were added and mixed. Upon the addition of UDP*G* (0.07 µM), samples were immediately vortexed and incubated for the required time intervals at 30°C. After each sample was incubated for exactly the desired duration (by keeping 15 s intervals between samples for the addition of radio-labelled substrate and then maintaining the same time gap to stop the reaction), reactions were terminated by the addition of 125 µl 0.45% (w/v) NaCl. Glycosides were extracted by adding 300 µl of H_2O saturated ethyl acetate to each reaction followed by vortexing for 10 s to separate the aqueous (bottom) and organic solvent (top) layers. 150 µl (of a total of 300 µl) of the extracted radiolabelled material was mixed with 4 ml Ecoscint liquid scintillation fluid (ND) in a 6 ml polyethylene vial (PE). Each sample was counted for 3 min by the Tricarb 2200 CA liquid scintillation counter (PE) against a stored quench curve prepared using commercially supplied standards (PE) with known quench values in each vial. The incorporation of UDP*G* was determined by subtracting counts present in control assays (incubations in the absence of acceptor), which were typically less than 100 dpm per assay using CFEs as well as purified UGTs.

Component	**Volume** (µl)	**Final amount/ concentration** (in 75 µl)
1-2 mM substrate (acceptor)		0-200 µM
50,000 dpm UDP*G* (0.07 µM, specific activity 302 mCi/mmol)		
100 mM TRIS-HCl; pH 8.0	50	67 mM
1 mg/ml purified enzyme in (5 mM TRIS-HCl; pH 8.0)		0.5 µg
18.2 MΩ/cm H_2O to a final volume of 75 µl		

Table 2.12 (a) Assay constituents for UGT activity using UDP*G* and purified UGTs.

Component	**Volume** (µl)	**Final amount/ concentration** (in 75 µl)
1 mM cholesterol (acceptor)	2.8	37 µM
50,000 dpm UDP*G* (0.07 µM, specific activity 302 mCi/mmol)		
CFE	60	
100 mM TRIS-HCl; pH 8.0 to a final volume of 75 µl		

Table 2.12 (b) Assay constituents for HTS for active UGTs using CFEs. Sample and control reactions were set up and incubated for 30 min at 30°C.

3. Cloning, expression and characterisation of UGTs

3.1 Introduction

Owing to its relatively smaller size, the *S. cerevisiae* genome was the first eukaryotic genome to be sequenced (Klionsky *et al*. 2003). Due to the presence of homologs of yeast genes in other eukaryotes, gene products of which are functional homologs in many cases, the role of yeast genes and gene products may be directly applied to other organisms.

The autophagy-related protein Atg26, ORF YLR189C from *S. cerevisiae* has been previously characterised as a USGT and was designated as UGT51 (Warnecke *et al.* 1999). UGT51 from *S. cerevisiae* (ScUGT51) is located on chromosome XII, which is approx. 1.1 Mb in length and consists of 534 ORFs (Johnston *et al*., 1997). **Fig. 3.1** shows the map of chromosome XII of *S. cerevisiae* highlighting the location of ScUGT51. Autophagy is a conserved process in eukaryotes and helps in nutrient recycling during starvation and stress as well as being involved in pathogen defence (Grille *et al.* 2010; Meijer *et al.* 2007). Although the homologs of ScUGT51 in *Pichia pastoris* and *Yarrowia lipolytica* have been shown to be involved in cellular degradation of proteins and pexophagy (Nazarko *et al.* 2007a; Nazarko *et al.* 2007b; Stasyk *et al.* 2003) recent studies have shown that ScUGT51 is not involved in such pathways and as such its role needs to be further elucidated (Cao and Klionsky 2007).

Warnecke *et al*. (1999) have experimentally shown that the majority of the *N*-terminal domain of ScUGT51 was not required for USGT activity. In addition to retaining its functional activity, the *N*-terminally truncated gene product was soluble and functionally stable when expressed in *E. coli*. The shortest *C*-terminal domain of ScUGT51 vital for enzyme function consists of sequence of 476 amino

acids resulting from a deletion of the first 722 *N*-terminal amino acids (Warnecke *et al.* 1999). The truncated form of ScUGT51 will hereby be designated as ScUGT51'.

This study has attempted to exploit the observation that ScUGT51' and related UGTs with *N*-terminal deletions could be available to detailed biochemical analysis as pure enzymes.

This chapter will describe the identification and characterisation of ScUGT51'and putative USGTs from related organisms based on primary sequence similarity with ScUGT51'.

In keeping with the nomenclature proposed for the yeast autophagy-related proteins (Klionsky *et al.* 2003), all UGTs described here will be denoted with a genus (capital) and species (lower case) prefix for example ScUGT51' for the *S. cerevisiae* UGT and the corresponding gene/gene fragments will be indicated in italics for example *ScUGT51*.

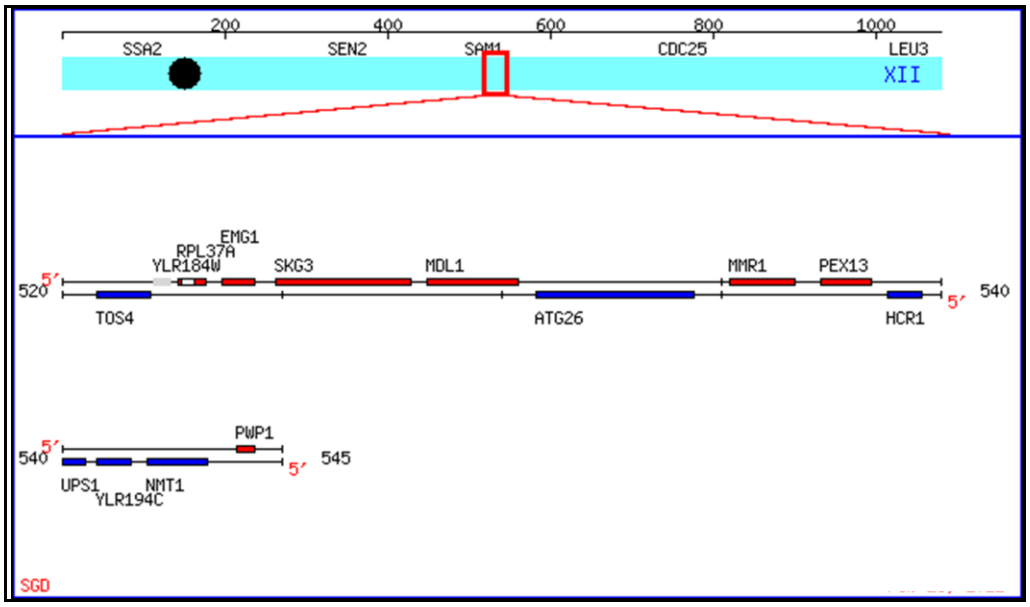


Fig. 3.1 Map of chromosome XII of *S. cerevisiae.* ScUGT51 (ATG26) is located on the Crick strand (blue) with chromosomal coordinates 534393-530797 bp. It is flanked by MDL1 and MMR1, inner and outer mitochondrial membrane proteins respectively on the Watson (red) strand. The black dot represents the location of the centromere. (Figure taken from the *Saccharomyces* genome database (SGD), Stanford University).

3.2 Materials and Methods

3.2.1 Bacterial and yeast strains

The required media and growth conditions of bacterial and yeast strains described in this chapter are detailed in section **2.2.**

3.2.2 Identification and bioinformatic analysis of putative UGT encoding ORFs

All identified ORFs, their corresponding amino acid sequence and other relevant information of the source organism and the genes of interest were obtained from the National Centre for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/).

Amino acid sequences of all ORFs were probed for the presence of signal peptides, trans-membrane spanning domains and probable sites for restriction endonucleases as follows:

(a) The amino acid sequence encoded by each ORF was searched for the presence of signal peptides using the Technical University of Denmark's Centre for Biological Sequence Analysis (CBS) web server, SignalP 3.0 (http://www.cbs.dtu.dk/services/SignalP/) (Emanuelsson *et al.* 2007).

(b) The amino acid sequences of each ORF were scanned for the presence of any trans-membrane helices (TMHs) using the CBS web server TMHMM, v.2.0 (http://www.cbs.dtu.dk/services/TMHMM/) and the Phobius tool (European Bioinformatics Institute (http://www.ebi.ac.uk/Tools/phobius/).

Protein localisation sites in the amino acid sequences were predicted and analysed using the PSORT Server (University of Tokyo, available at http://psort.ims.u-tokyo.ac.jp/).

(c) The deduced gene fragments were checked for the presence of internal restriction sites for *Xba*I and *Xho*I using the NEB Web Cutter 2.0 program (http://tools.neb.com/NEBcutter2/).

ClustalW web program (http://www.ebi.ac.uk/Tools/msa/clustalw2/) was used to determine primary sequence alignments. The Pfam database (http://pfam.sanger.ac.uk/) was used to determine organisation of UGT domains.

3.2.3 PCR amplification and LIC-cloning of gene fragments

Full length nucleotide and amino acid sequences of gene fragments as well as the map and cloning region of the pET-28a derived LIC vector are included in Appendix **A.**

3.2.3.1 PCR for gene fragments

Two primers were designed for each ORF under investigation, a forward primer and a reverse primer. These were designed using the Web primer program (http://www.yeastgenome.org/cgi-bin/web-primer). Primers were selected such that each primer was a minimum of 18 nucleotides in length and that the difference in T_m (DNA melting temperature) values of the forward and reverse primers was kept to a maximum of 5^oC.

For all ORFs, the forward primer is identical to the first 19-21 nucleotides in the 5'→3' direction on the 'sense' strand. Reverse primers designed are reverse complimentary to the last 18-22 bases on the 'sense' strand. To enable cloning of the PCR products into the LIC vector, LIC-specific ends (**Table 3.1** in bold) were added to the 5' end of each of the forward and reverse primers. The T_m values were then calculated for each of the primers including and excluding the LIC-specific sequences using the following equation and are recorded in **Table 3.2 (a)**.

T_m= 69.3+ 0.41 x GC% - (650 ÷ total length of sequence)

Primers were paired together for the PCR resulting in an amplified gene that was annealed to the LIC vector such that an *N*-terminally hexahistidine-tagged protein could be expressed.

The calculated and optimised primary and secondary annealing temperatures for PCR are summarised in **Table 3.2 (b).**

ORF	Forward Primer	Reverse Primer
Sc	5'-**caccaccaccac**ATGATTGATGAGAATCCGCAC-3'	5'-**gaggagaaggcgcg**TTAAATCATCGTCCACCCTTCA-3'
Pp	5'-**caccaccaccacatg**AACCAGTATTCTACGTCAGA-3'	5'-**gaggagaaggcgcg**TTACACTTCAAACCATGATC-3'
Pg	5'-**caccaccaccacatg**TTAGAAGATTCACCACTCATC-3'	5'-**gaggagaaggcgcg**TTAGTCGTTAGTTGTCGTTG-3'
Kp	5'-**caccaccaccacatg**GTTGAGGATAATCCATATTTT-3'	5'-**gaggagaaggcgcgtt**AGACCATGCTGGTGAGCG-3'
Pa	5'-**caccaccaccacatg**ATAGAAGAGCATCCATTGACC-3'	5'-**gaggagaaggcgcgtt**AAACCAGTAACCACGAGCC-3'
Kl	5'-**caccaccaccacatg**GTTGAAAAAAACCCGTATT-3'	5'-**gaggagaaggcgcg**TTATATCAGAAGCCAGGATC-3'
Dh	5'-**caccaccaccacatg**TTAGAAGATTCTCCATTCTAT-3'	5'-**gaggagaaggcgcgtta**TTATTTGTTAGCATCTGTAAC-3'

Table 3.1 Primer pairs for the PCR amplification and cloning of probable UGTs into the LIC vector. LIC-specific ends are shown in bold and lowercase.

ORF source organism	Tm of FP excluding LIC sequences in calculation (^{o}C)	Tm of RP excluding LIC sequences in calculation (^{o}C)	Tm of FP including LIC sequences in calculation (^{o}C)	Tm of RP including LIC sequences in calculation (^{o}C)
Sc	53.0	55.0	70.7	72.9
Pp	53.2	51.5	70.6	70.7
Pp	53.9	53.2	70.6	71.9
Kp	50.0	56.6	68.3	75.0
Pa	55.9	56.7	71.7	74.2
Kl	48.0	53.2	68.3	71.9
Dh	50.0	50.0	75.0	72.8

Table 3.2 (a) Calculated T_m values of primers including and excluding LIC specific sequences in calculation. (FP- forward primer; RP-reverse primer)

ORF source organism	Primary annealing temperature (^{o}C)		Secondary annealing temperature (^{o}C)	
	C	O	C	O
Sc	48.0	54.5	65.7	77.5
Pp	46.0	51.0	65.5	72.4
Pg	48.2	48.0	65.6	69.8
Kp	45.0	48.5	63.3	71.5
Pa	50.9	42.7	66.7	65.7
Kl	43.0	48.5	63.3	71.5
Dh	45.0	45.5	63.3	68.5

Table 3.2 (b) Calculated (C) and optimised (O) primary and secondary annealing temperatures for paired primers. Temperatures were calculated by subtracting 5^{o}C from the lower T_m value of primer pair (**Table 3.2 (a)**). Optimised temperature pair was used for PCR.

A 1:5 dilution of the primers (100 μM) was prepared in 18.2 MΩ/cm H_2O for the PCRs which were set up as follows:

All components were pulsed by centrifugation for 5 s prior to use and were added to a PCR tube as shown in **Table 3.3 (a)**. KOD Hot Start DNA polymerase was added in the end and was mixed by swirling the pipette tip in solution. The PCR tubes were then pulsed by centrifugation for 5 s.

Component	**Volume** (μl)	**Final concentration** (in 50 μl reaction)
1.0 U/μl KOD Hot start DNA polymerase	1	1 U
10 x buffer for KOD Hot start DNA polymerase	5	1 x
Genomic template DNA (yeast genomic DNA- 0.1 μg/μl; LIC vector template- 1 ng/μl.)	0.5	1 ng (yeast genomic DNA); 0.01 ng/μl (LIC vector template)
20 mM Forward primer	1	0.4 mM
20 mM Reverse primer	1	0.4 mM
2 mM each dNTPs	5	0.2 mM (each)
25 mM $MgSO_4$	2	1 mM
18.2 MΩ/cm H_2O	34.5	

Table 3.3 (a) PCR components for amplification of DNA fragments and LIC vector.

The PCR cycle used to amplify DNA fragments is tabulated in **Table 3.3 (b)** using the annealing temperatures described in **Table 3.6.3 (c)**. KOD Hot Start DNA polymerase was used for PCR product amplification. 5 μl of each PCR product was mixed with 1 μl loading dye (section **3.1.3.3**) and analysed by AGE. The samples were run with a DNA size standard. The target gene fragment was expected to be located at approx. 1431 base pairs (bp) for Sc; 1395 bp for Pp.; 1602 bp for Pg; 1434 bp for Kp.; 1413 bp for Pa.; 1473 bp for Kl and 1632 bp for Dh.

Step	Temp (^{o}C)	Time (h:min:s)	Purpose
1	95.0	00:02:00	Initial denaturation
2	95.0	00:00:15	Denaturation
3	**X**	00:00:30	Annealing
4	72.0	00:01:38	Elongation
5			Step 2- 4: repeated 4 times = 5 cycles
6	95.0	00:00:15	Denaturation
7	**Y**	00:00:30	Annealing
8	72.0	00:01:38	Elongation
9			Step 6-8, repeated 24 times= 25 cycles
10	72.0	00:10:00	Final Extension
11	10.0	Hold	Storage

Table 3.3 (b) PCR program used for the amplification of ORFs. X= Primary annealing temperature for primers (as indicated in **Table 3.2 (b); Y**= Secondary annealing temperature (as indicated in **Table 3.2 (b)**.

3.2.3.2 PCR amplification of LIC vector followed by *Dpn*I digestion

LIC vector sequence was amplified via PCR (See **Table 3.3 (a)** for PCR components) and the PCR cycle used for amplification is described in **Table 3.3 (c)**. 5 µl of the PCR product was mixed with 1 µl of loading dye and was analysed by AGE. 1 µl of *Dpn*I (20,000 U/ml, New England Biolabs) was then added to 45

µl of the PCR product along with 5 µl of 10 x Buffer 2 (recipe in **Table 2.6**) and incubated in a H_2O bath at 37^oC for 1 h.

Target gene and LIC vector concentration was determined using the Bio-Rad DNA quantification software (Quantity One) using a DNA size standard (Hyper Ladder 1) with known amounts of DNA in each fragment.

Step	Temp (^oC)	Time (h:min:s)	Purpose
1	95.0	00:02:00	Initial denaturation
2	95.0	00:00:30	Denaturation
3	75.0	00:00:30	Annealing
4	72.0	00:06:00	Elongation
5			Step 2-4, repeated 34 times
6	72.0	00:10:00	Final Extension
7	10.0	Hold	Storage

Table 3.3 (c) PCR program used for the amplification of LIC vector

3.2.3.3 Ligation-independent cloning (LIC)

3.2.3.3.1 T4 polymerase (LIC T4 pol) reaction for PCR products

To enable ligation independent cloning (Bonsor *et al*., 2006), PCR products of amplified target inserts and LIC vector (section **2.2.1.13**) were treated with LIC-qualified T4 polymerase. Reaction components for LIC vector and target inserts are tabulated **Table 3.4 (a)** and **(b)**.

Component	**Volume (**µl)	**Final amount/ concentration** (in 200 µl)
Linearised LIC vector DNA (Section **2.2.1.13**) 25 ng/µl	45	1.1 µg
10x T4 polymerase buffer	20	1x
25 mM dTTP	20	2.5 mM
100 mM DTT	10	5 mM
2.5 U/µl T4 pol (LIC-qual)	4	10 U
18.2 MΩ/cm H_2O to a final volume of 200 µl	101	

Table 3.4 (a) T4 polymerase reaction of LIC vector.

Component	**Volume** (µl)	**Final amount/ concentration** (in 20 µl)
PCR product (insert) (10-100 ng/µl) (Section **2.2.1.13**)	14.6	0.2 pmol
10x T4 polymerase buffer	2	1x
25 mM dATP	2	2.5 mM
100 mM DTT	1	5 mM
2.5 U/µl T4 polymerase (LIC-qual)	0.4	1 U

Table 3.4 (b) T4 pol treatment of target inserts

For each reaction, the reaction constituents were mixed gently with a pipette tip and the reaction mixture was incubated at 22^{o}C for 30 min followed by further

incubation at 75^{o}C for 20 min. Following LIC T4 pol treatment, LIC vector was purified as described section **2.2.1.6**.

3.2.3.3.2 Nucleotide removal following T4 polymerase (T4-pol) reaction

This procedure was carried out for nucleotide removal following LIC-T4 pol reaction of LIC vector (section **2.2.1.14**). This was achieved using a kit (**Table 2.7**) and the following method: All centrifugation steps were performed at 10,000 x *g*. 5 volumes of Buffer PN (recipe undisclosed by manufacturer) was added to 1 volume the sample and mixed. A spin column was placed in the 2 ml collection tube and 600 µl of the sample applied to the spin column and centrifuged for 1 min to bind DNA. The flow-through was discarded and the spin column placed back into the same collection tube. Note: for samples > 600 µl, the same spin column was re-loaded with the remaining mixture and centrifugation repeated. To wash the column, 750 µl of buffer PE (recipe undisclosed by manufacturer) was added and centrifuged for 1 min. The flow-through was discarded and the spin column was placed back in the same collection tube and centrifuged for an additional 1 min. The spin column was then placed in a sterile 1.5 ml microcentrifuge tube. To elute DNA, 100µl of sterile 18.2 MΩ/cm H_2O (pH 8.0) was added to the centre of the column membrane and the column was centrifuged for 1 min. This step was repeated and the resulting 200 µl of eluant was stored at -20°C until further use.

3.2.3.3.3 Ligation of target inserts into LIC vector followed by transformation in cloning host *E. coli* TOP10

Each of the T4 pol treated reactions (2 µl; 50-100 ng/µl) was added to 1 µl of the T4 pol treated LIC vector (192 ng/µl) and the reaction mixture incubated at room temperature for 10 min. 1 µl of 100 mM EDTA was then added to the reaction mixture and mixed with a pipette tip and the solution incubated at room temperature for further 10 min. To check *E. coli* transformation efficiency, 2 µl pET-28a vector DNA (0.01 ng/µl) was used as a control reaction.

4 µl of each of the LIC annealing products were transformed into to 50 µl of freshly prepared *E. coli* TOP 10 chemically competent cells as described in section **2.2.1.12.**

3.2.4 In-gel trypsin digestion and identification of r-proteins by peptide mass fingerprinting

Protein bands of interest were excised from polyacrylamide gels following SDS-PAGE (section **2.2.2.1**) and were cut into 1 mm x 1 mm slices using a sterile scalpel. These were then transferred to clean siliconized 0.5 ml tubes and could be stored at -20°C until future use. All incubation steps were performed at room temperature unless stated otherwise. Gel slices were washed three times by adding 100 µl of 100 mM NH_4HCO_3, pH 8.0 (section **2.1.7**) and 60 µl ACN and left orbitally shaking for 30 min. The liquid was discarded in between each wash step. Following this, the gel slices were dehydrated by adding 100 µl of ACN and incubated for 5 min. This step was repeated following which gel slices were dried

by centrifugal evaporation for 15 min. The recovered peptides were digested with 25 µl of 0.1 µg/µl trypsin (section **2.1.7**) in 50 mM NH_4HCO_3 on ice for 30 min. 30 µl of 50 mM NH_4HCO_3 was added to the tubes prior to incubation at 37 °C overnight. The peptides were extracted by incubating in 30 µl of a 50% (v/v) ACN with 5% (v/v) formic acid solution and the tubes were left shaking for 30 min. The solution containing the digested peptides was transferred to a clean 1.5 ml microcentrifuge tube. A 30 µl solution of 83% (v/v) ACN with 0.1% (v/v) formic acid was added to the original tube with gel slices followed by incubation for another 30 min. The liquid containing the leftover digested peptides (if any) was pooled with the initial collection of digested peptides. For protein identification, the samples were freeze-dried, dissolved in 10 µl of start buffer A (section **2.1.6.2**) and the solution transferred into a microtitre well plate. Samples were analysed by an HCTUltra MS (Bruker Daltonics) via an electron spray ionisation (ESI) source. Instruments were controlled and the MS-MS spectrum obtained using HyStar™ 3.2 software (Bruker Daltonics). Peptide analysis was performed at a flow rate of 3 µl/min with a nebulizer pressure of 10 psi and a drying gas flow rate of 5.0 L/min at a drying gas temperature of 300°C.

Database searches were performed against NCBI non-redundant (nr) peptide database (http://ncbi.nlm.nih.gov/BLAST/) using the Mascot MS/MS ions search program (Matrix Sciences, UK; http://matrixscience.com/search_form_select.html).

Preliminary qualitative UGT activity assays and TLC analysis were performed based on assay conditions stated in Madina *et al*., 2007 and Warnecke *et al*., 1999. Acceptor and donor substrates described in this chapter are listed in **Table 2.8 (a)**

and **(b)**. Stocks of all acceptor aglycones were prepared in ethanol and made fresh on the day of use. UDPG/UDPGA stocks were prepared in 18.2 MΩ/cm H_2O and stored at -20°C. These were diluted and kept on ice prior to assaying. The assay mixture constituents are summarised in **Table 3.5**. Reaction mixtures were scaled up if required (to a total reaction volumes of 200-300 µl) for HPLC separation and MS analysis as described in section **2.2.4**.

Component	**Volume** (µl)	**Final amount/ concentration/amount** (in 100 µl)
4 mM substrate (acceptor)	20	0.8 mM
40 mM UDP-sugar (donor)	2	0.8 mM
100 mM TRIS-HCl; pH 8.0	10	10 mM
1 mg/ml purified enzyme in (5 mM TRIS-HCl; pH 8.0)		2 µg
18.2 MΩ/cm H_2O to a final volume of 100 µl		

Table 3.5 Standard UGT activity assay constituents.

All reactions were set in triplicate along with control reactions where 2 µl of UDP-sugar donor was substituted with 2 µl 18.2 MΩ/cm H_2O. Reactions were also set up with 1M $MnCl_2$ in the reaction buffer (0.1 mM final concentration).

For extraction of lipid products, reactions were incubated at 30°C for 2 h and were terminated by the addition of 0.9 ml of 0.45% (w/v) NaCl solution and 4 ml of chloroform: methanol, 2:1 v/v. The reactions were vortexed for 20 s after which the aqueous phase (top layer) was discarded and the organic phase (bottom layer) containing the reaction product was collected and dried at room temperature using a sample concentrator with an external N_2 gas supply. The samples were re-constituted in 30 µl chloroform: methanol, 2:1 (v/v) and spotted on a TLC plate (silica gel 60, F_{254}) and run using a mixture of 85:15 (v/v) chloroform: methanol as the mobile phase. For lipid detection, plates were developed with spray of 15% (w/v) phosphotungstic acid (SIG) in ethanol and visualised by charring using a heat gun. Samples were analysed against an authentic testosterone glycoside standard (ProStrakan).

3.2.6 HPLC-MS analysis of reaction products

HPLC-MS analysis of standard assay reaction products was performed on a Luna CN Phenomenex column (5.0 cm x 2 mm 5µm) using a Thermo Finnigan Surveyor with analytes detected by an ion-trap mass-spectrometer (Thermo Finnigan LCQ Advantage). MS detection was performed in selected ion monitoring (SIM) mode with an ion spray voltage of 4.5 kV and capillary temperature of 230°C. 100 µl of post assay reactions (section **2.2.3.1**) after 2-3 h incubation periods were terminated by the addition of 100 µl 0.3 M HCl and were mixed with 800 µl distilled H_2O prior to loading using an autosampler at an injection volume of 20 µl and flow rate of 150 µl/min. The eluant was 30% (v/v) acetonitrile, 1% (v/v) acetic acid used in an isocratic mode with a run time of 5

min. Glycosides were identified using mass spectral data visualised using the Xcalibur software suite. The identity of the testosterone glycoside was confirmed by comparison with an authentic standard kindly provided by ProStrakan (Galashiels, UK).

3.2.7 Kinetic analysis of UGTs

Kinetic parameters (V_{max}, K_M and K_{cat}) of UGTs against different acceptor substrates were determined using UDP*G* as the donor substrate. For all kinetic analysis, 4 - 5 substrate concentrations (determined empirically) in the limiting range were chosen and reactions were performed as detailed in section **2.2.3**. Michaelis-Menten and Lineweaver-Burk plots were prepared using GraphPad Prism software version 5.04 for Windows (GraphPad Software, San Diego California USA). Specific activities (SpA) of UGTs were determined under standard assay conditions by measurement of an initial reaction rate at saturating substrate concentration.

Raw dpm data was treated as follows:

Data from LSC was expressed in dpm as described in section **2.2.3.** Amount of product (sterol glycoside) formed in pmol was calculated as follows:

a) Measured dpm (x) was converted to curies (Ci) as 1Ci= 2.22 x 10^{12} dpm; therefore dpm value/ 2.22 x 10^{12} = y Ci
b) Amount of radiolabelled product formed (y) was then converted to amount (in mmol)

y/specific activity of UDP*G* (0.302 Ci/mmol) = *z* mmol

$z \times 10^9$ = Amount of product formed (pmol).

Therefore, $x \times 10^9 / 0.302 \times 2.22 \times 10^{12} = x \times 670.4$

3.3 Results

The molecular weight and extinction coefficient values of proteins were predicted based on the identified amino acid sequences including the *N*-terminal hexahistidine tag. These analyses were performed using the ProtParam tool within the web program ExPASy (expert protein analysis system; Swiss Institute of Bioinformatics, http://expasy.org/tools/protparam.html).

The nucleotide sequences corresponding to each truncated amino acid sequence were then searched to check for the presence of restriction sites of restriction enzymes *Xba*I and *Xho*I using the NEB v2.0 web program (NEBcutter, http://tools.neb.com/NEBcutter2/).

3.3.1 Identification and investigation of ORFs

A Pfam database (http://pfam.sanger.ac.uk/) analysis (Finn *et al.* 2008) of the ScUGT51 sequence revealed the presence of the pleckstrin homology (PH) domain (PF00169), GRAM domain (PF02893), a GT family 28 (GT28) domain and a UDP-glucoronosyl/UGT (UDPGT) domain (PF00201) (**Fig. 3.2 A**.). Warnecke *et al*. (1999) showed by successive deletions of the *N*-terminus of ScUGT51 followed by *in vitro* analysis that only the GT28 and UDPGT *C*-terminal domains was required for the synthesis of several steryl glycosides including the natural ergosteryl-β-glucoside. The shortest, truncated functional UGT sequence, ScUGT51' consisted of 476 amino acids following the deletion of 722 *N*-terminal amino acids.

In an attempt to identify and characterise novel UGTs on the basis of sequence similarity approaches and to analyse the function of conserved residues and motifs, a BlastP (Altschul *et al*.,1997) search was performed using ScUGT51' as a template to probe the sequenced protein database. At the time of conducting the BLAST search at the beginning of this study, ScUGT51' displayed a high percentage of overall sequence identity, ranging between 63-69% with the characterised UGTs isolated from other yeasts such as the UGT51B1 of *P. pastoris* (Pp), UGT51 from *Candida albicans*, a putative USGT from *Yarrowia lipolytica* as well as protein products Atg26p from *Pichia angusta* (Pa) and AFR681Wp from *Ashbya gossypii* and uncharacterised protein products Kpol_2000p77 from *Kluveromyces polysporus* (Kp), 'unnamed' product from *Kluveromyces lactis* (Kl), PGUG_04770 from *Pichia guilliermondii* (Pg) and DEHA0E25168g from *Debaromyces hansenii* (Dh). However, the pairwise sequence identities of the non-truncated sequences as compared with ScUGT51 were between 33-45%, implying that the majority of the differences existed as a result of the variations within the *N*-terminal regions. This is further indicated by the organisation of the *N*-terminal PH and GRAM domains differs in all the identified sequences (**Fig. 3.2 B.**).

Among the first ten hits that displayed significant sequence identities, the sequences that were selected for further investigation in this study, in addition to the ScUGT51' were those of Pp, Pg, Pa, Kl, Kp and Dh (**Table 3.6**). Among these, only UGT51B1 from Pp had been previously characterised as a USGT (Warnecke *et al.* 1999).

An amino acid sequence alignment was performed using the ClustalW2 web alignment program (European Bioinformatics institute, http://www.ebi.ac.uk/Tools/msa/clustalw2/) with ScUGT51' and the selected sequences identified using Blastp searches. This allowed identification of the exact regions of all sequences that aligned with ScUGT51'. Following this, truncated versions of all identified sequences with appropriate *N*-terminal amino acid deletions made, were produced. **Fig. 3.2 C**. shows the sequence alignment of ScUGT51'with the BLAST-identified sequences. Regions of highly conserved amino acids were noted in all sequences (indicated by black boxes). In addition, all sequences also contain the USS (underlined), a characteristic feature of the GT-B fold UGT superfamily (Hu and Walker 2002). The presence of highly conserved residues in the *C*-terminal domain across all identified sequences strongly suggested that these could be functional homologues.

All deduced sequences were checked for the presence of *N*-terminal putative signal peptide and trans-membrane helices in order to avoid improper folding of the expressed proteins. This is because expressing proteins with a signal peptide could lead to the formation of insoluble aggregates which prevents them being studied in detail (Simonen and Palva, 1993). As expected, none of the selected sequences indicated the presence of TMHs or signal peptides.

Organism	GenBank Accession Nos. of UGTs	Reference
Saccharomyces cerevisiae (Sc) strain S288C	NP_013290	Johnston *et al*., 1997
Pichia pastoris (Pp) strain GS115	AAD29570	Warnecke *et al*., 1999
Pichia guilliermondii (Pg)	XP_001482815	Birren *et al*., unpublished data
Kluyveromyces polysporus (Kp)	XP_001646967	Scannell *et al*., 2007
Pichia angusta (Pa)	ABO31066	Meijer *et al*., 2007
Kluyveromyces lactis (Kl)	XP_452287	Dujon *et al*., 2004
Debaromyces hansenii (Dh)	XP_460322	Dujon *et al*., 2004

Table 3.6 Summary of yeast host strains used in this study.

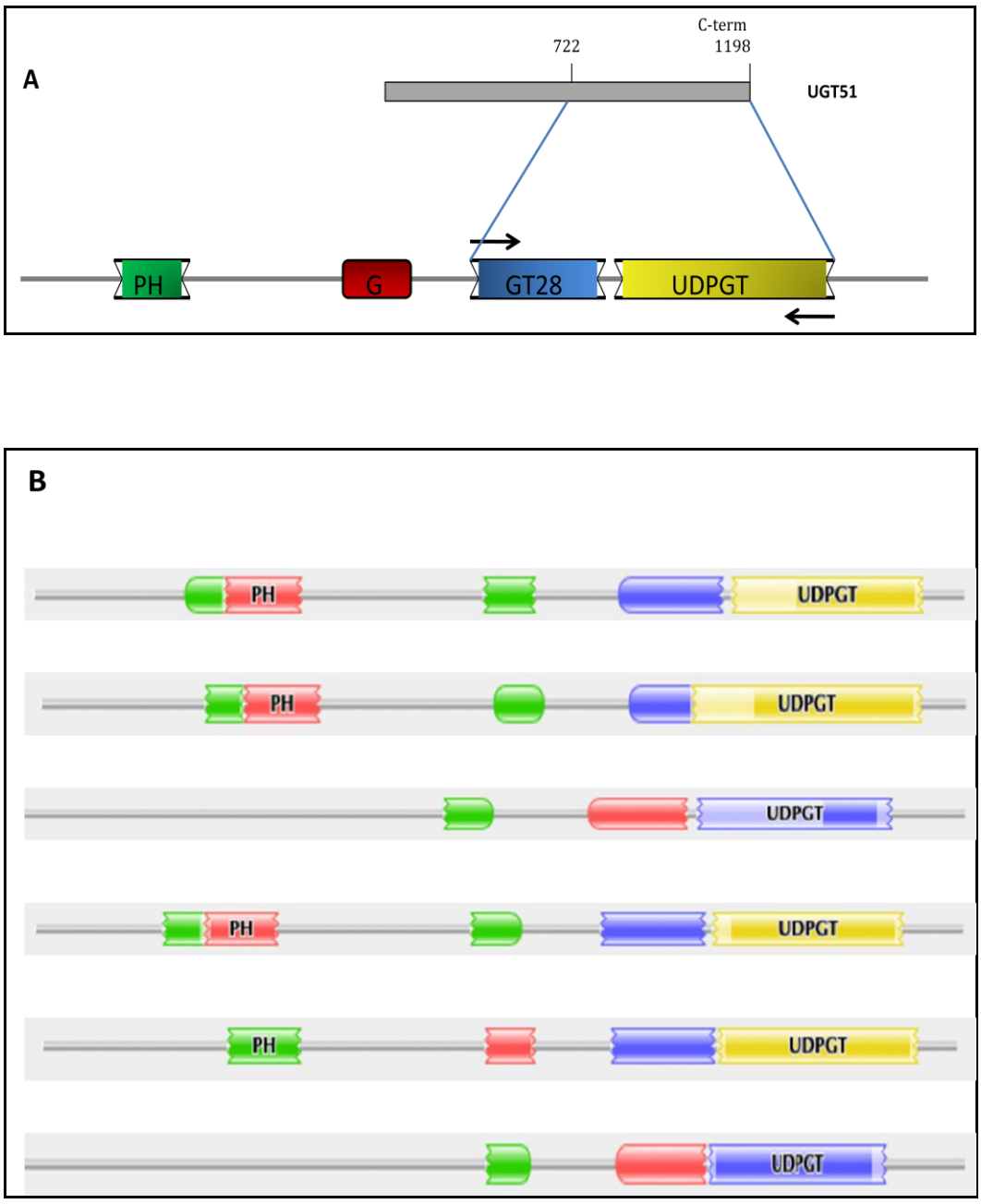


Fig. 3.2 Schematic showing overall organisation of UGT domains. A. Domains of ScUGT51**.** The extent *N*-terminal truncation of ScUGT51 is indicated by arrows and was conserved in all sequences. ScUGT51 consists of the PH and GRAM domains (G) involved in protein localisation; a GT family 28 sequence and the UDPGT domain; (Distances are not drawn to scale). **B**. PH, GRAM, GT28 and UDPGT domains of Pp, Pa, Pg, Kl, Kp and Dh (top to bottom) (Figures taken from Pfam database (Finn *et al*., 2008).

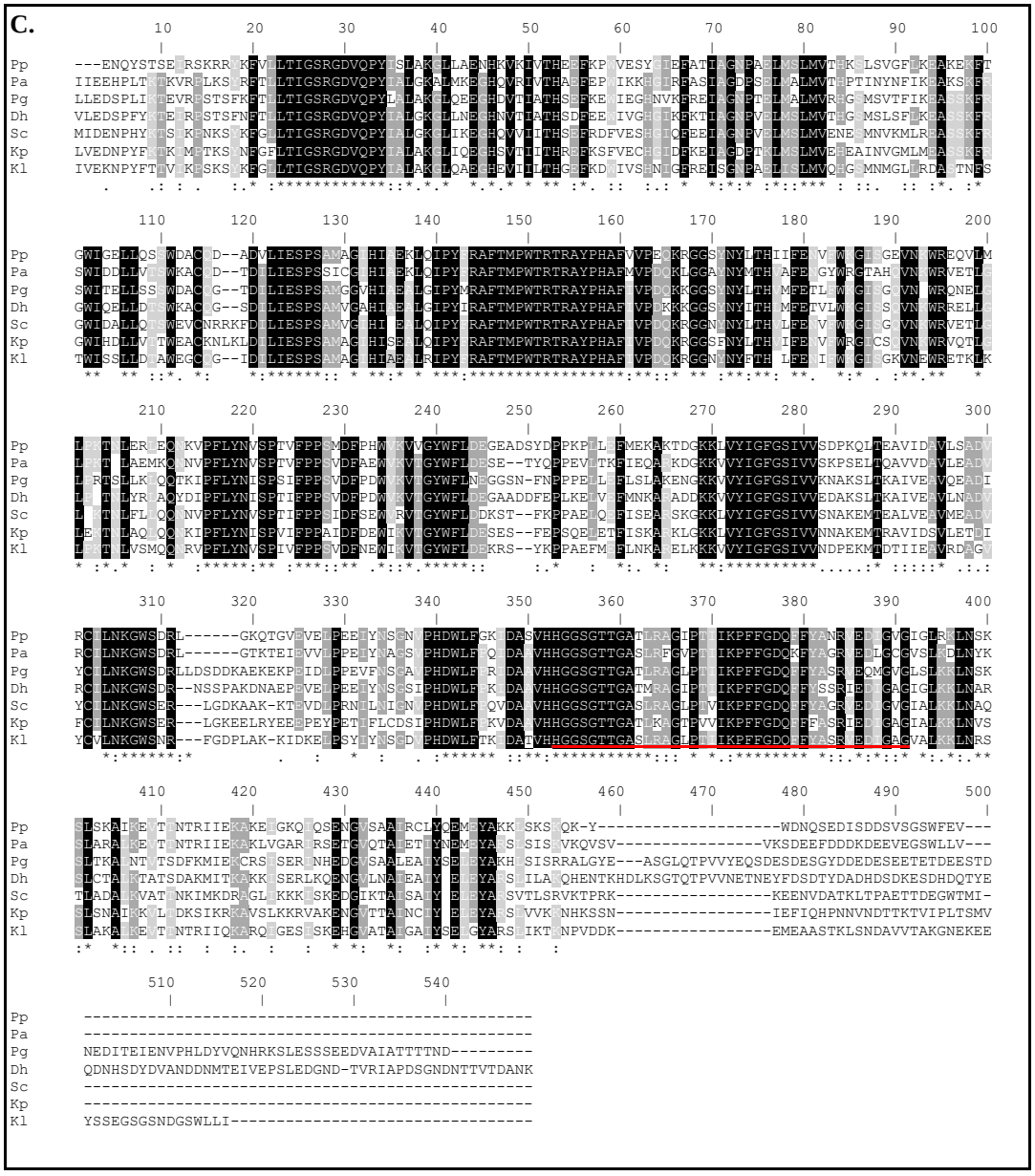


Fig. 3.2 C. ClustalW 2.0.12 alignment of ScUGT51' with deduced yeast sequences. Black boxes indicate identical amino acids in all 7 sequences. Grey boxes indicate identical amino acids in 5 or 6 sequences; the USS is underlined in red. The first amino acid residue of ScUGT51' shown here corresponds to the 723rd residue in the ScUGT51 and aligns with residues 748, 772, 1067, 1032, 741 and 491 in the deduced sequences of Pp, Pa, Pg, Dh, Kp and Kl respectively.

3.3.2 Cloning, UGT expression and purification

The ORFs of the gene fragments of interest were amplified by PCR and cloned into the LIC vector. The LIC vector is a modification of the pET-28a vector where the gene of interest can be inserted into the vector alleviating the need for time consuming and unreliable ligation steps for cloning (Bonsor *et al.* 2006). The LIC vector also encodes a hexa-histidine tag that can be incorporated at the *N*-terminus of the protein of the encoded gene thereby allowing one-step r-protein purification using IMAC.

3.3.2.1 PCR amplification of target genes followed by LIC

ORFs of gene fragments hypothesised to encode putative UGTs were amplified by PCR from chromosomal DNA extracted from the respective yeast strains (See **Fig. 3.3** for visualisation of DNA by AGE). Forward (F) and reverse (R) primer pairs that included LIC specific ends (section **3.2.3.1**) were used to generate PCR fragments (**Fig. 3.4 A-C**) corresponding to polypeptides with *N*-terminal deletions as described in section **3.1.1**. Optimised annealing temperatures for PCR are listed in **Table 3.2 (b)**. In all cases except for *P. guilliermondii*, it was observed that increasing the initial primary and consequently secondary annealing temperatures (**Table 3.7**) contributed to specific primer annealing thereby generating a cleaner target gene product.

LIC vector was prepared by PCR amplification where at the optimised annealing temperature of 75°C, a dense PCR band is observed (**Fig. 3.4 D**) between 5 Kb (Kilo bases) and 6 Kb when compared to the DNA size standard (L). Since the

size of linear LIC vector was expected to be approx. 5.4 Kb (See **Appendix A**), these results indicate the successful PCR amplification of the LIC vector. PCR amplified LIC vector was treated with *Dpn*I in order to digest the original circularised vector DNA template and prevent it from being preferentially transformed in *E. coli*.

Following PCR cleanup (section **2.2.1.6**), PCR products and LIC vector were treated with T4 pol (section **3.2.3.3.1**), that cleaved DNA fragments to result in the formation of 12-15 base 5'→3' specific non-complementary single-stranded overhangs to enable LIC. Treated products were annealed and transformed into chemically competent *E. coli* TOP 10 as described in section **2.2.1.12.** A transformation efficiency of 50-60 CFUs per 192 ng LIC vector was achieved compared to the pET-28a positive control where 270 CFUs per 0.02 ng vector (expected 100-200 CFU) were noted.

ORF of source organism	**1**		**2**		**3**		**4**	
	P	**S**	**P**	**S**	**P**	**S**	**P**	**S**
Kp	28.9	51.9	38.2	61.2	48.5	71.5		
Pa	33.1	56.1	42.7	65.7	53.4	76.4		
Sc	35.6	58.6	45.5	65.5	54.5	77.5		
Kl	28.9	51.9	38.2	61.2	48.5	71.5		
Dh	28.9	51.9	35.6	58.6	45.5	68.5		
Pp	28.9	51.9	35.6	58.6	45.5	68.5	51.0	72.4
Pg	40.7	63.3	48.0	69.8	45.2	67.3		

Table 3.7 Optimisation of (P) and secondary (S) annealing temperatures (^{o}C) for PCR. The corresponding results are shown in **Fig. 3.4** (**A-C**).

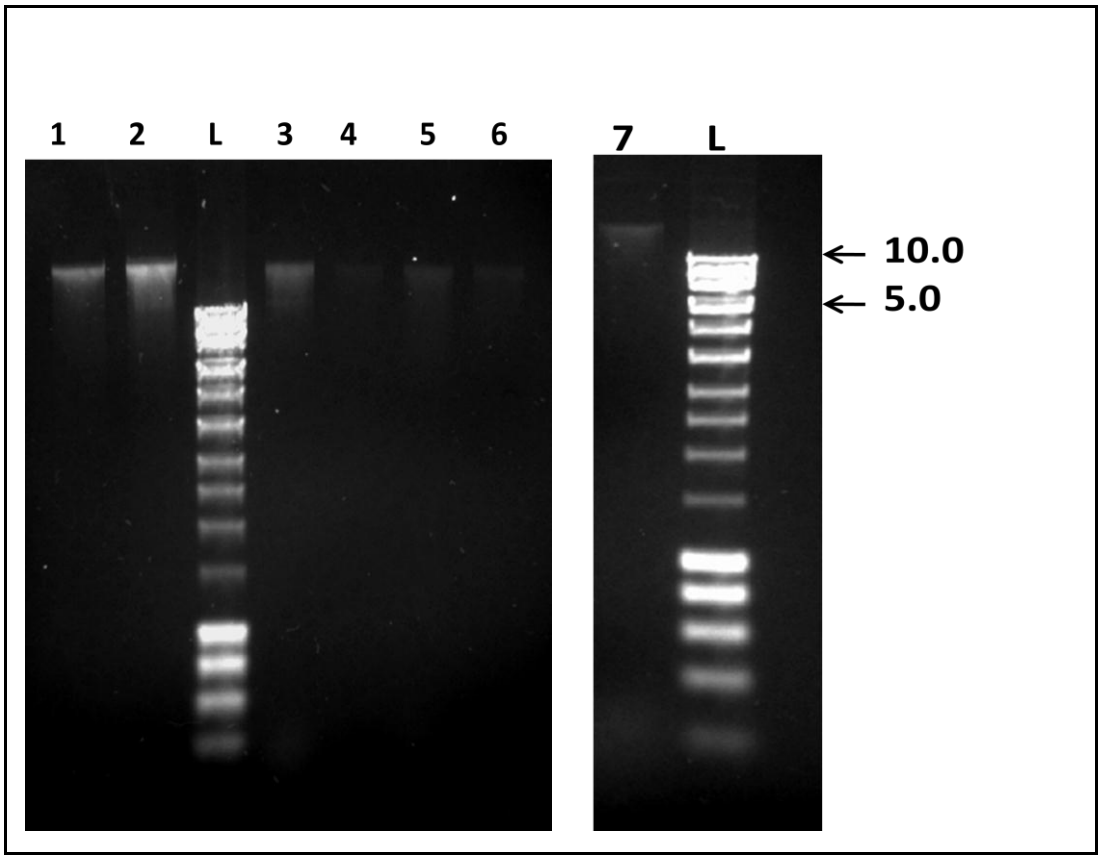
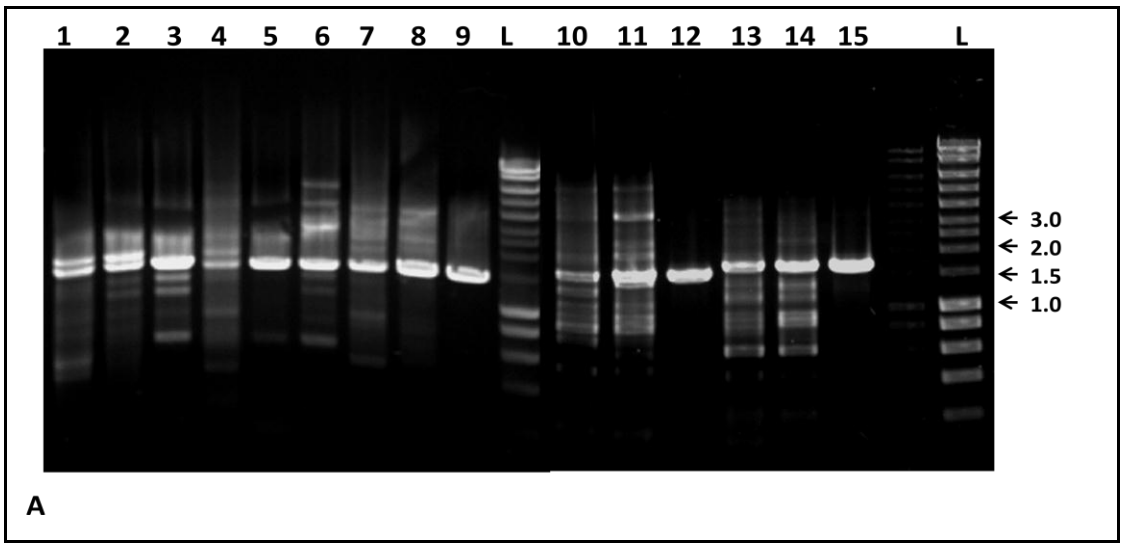


Fig. 3.3 Visualisation of genomic DNA of yeasts following AGE. Sc, Pp, Pa, Pg, Kl, Dh, Kp (Lanes 1-7). L= m.wt DNA size standards ranging between 200 bp and 10 Kb.



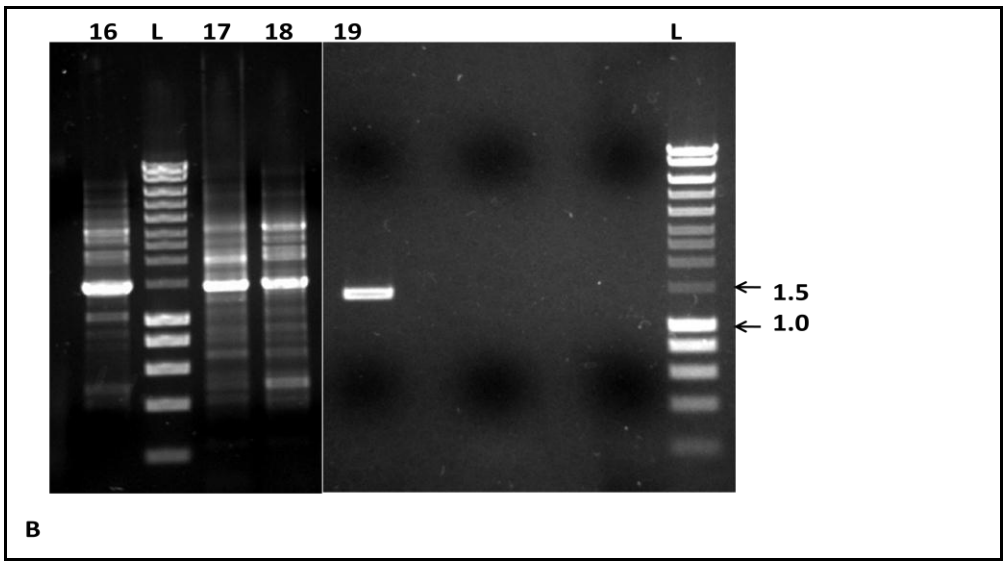


Fig. 3.4 Visualisation of gene fragments following PCR. Annealing temperature sets used are listed in **Table 3.7**. **A.** *Lanes 1-3*: PCR products for Kp at annealing temperature set 1-3 (expected size 1434 bp + LIC ends); *Lanes 4-6*: PCR products for Pa at annealing temperature set 1-3 (expected size 1413 bp +LIC ends); *Lanes 7-9*: PCR products for Sc at annealing temperature set 1-3 (expected size 1431 bp + LIC ends); *Lanes 10-12*: PCR products for Kl at annealing temperature set 1-3 (expected size 1473 bp + LIC ends); *Lanes 13-15*: PCR products for Dh using annealing temperature set 1-3 (expected size 1632 + LIC ends); **B.** *Lanes 16-19*: PCR products for Pp at annealing temperature set 1-3 (expected size 1395+ LIC ends). L= DNA size standards ranging between 200 bp and 10 Kb.

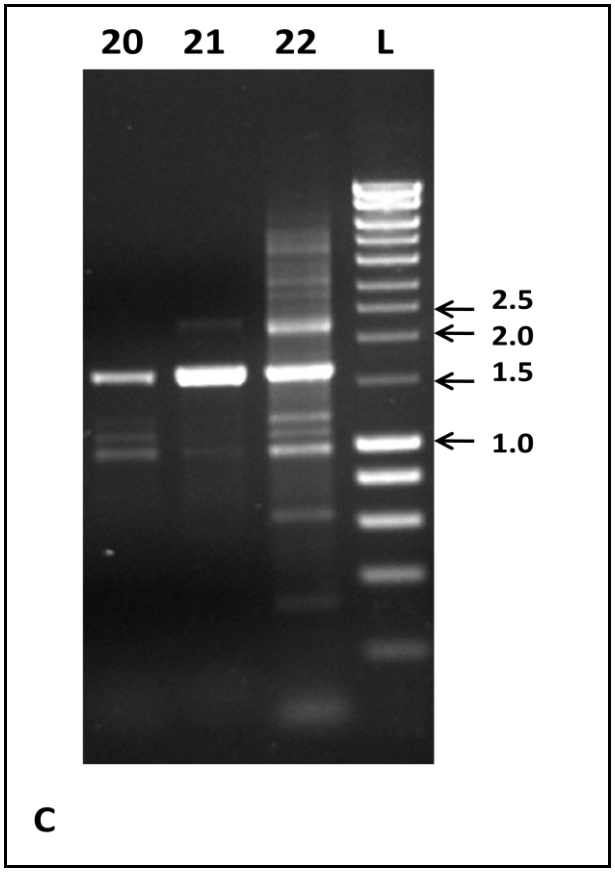


Fig. 3.4 C. Visualisation of gene fragments following PCR. Annealing temperature sets used are listed in **Table 3.7** *Lanes 20-22*: PCR products for Pg using annealing temperature set 1-3 (expected size 1602 bp + LIC ends). L= DNA size standards ranging between 200 bp and 10 Kb.

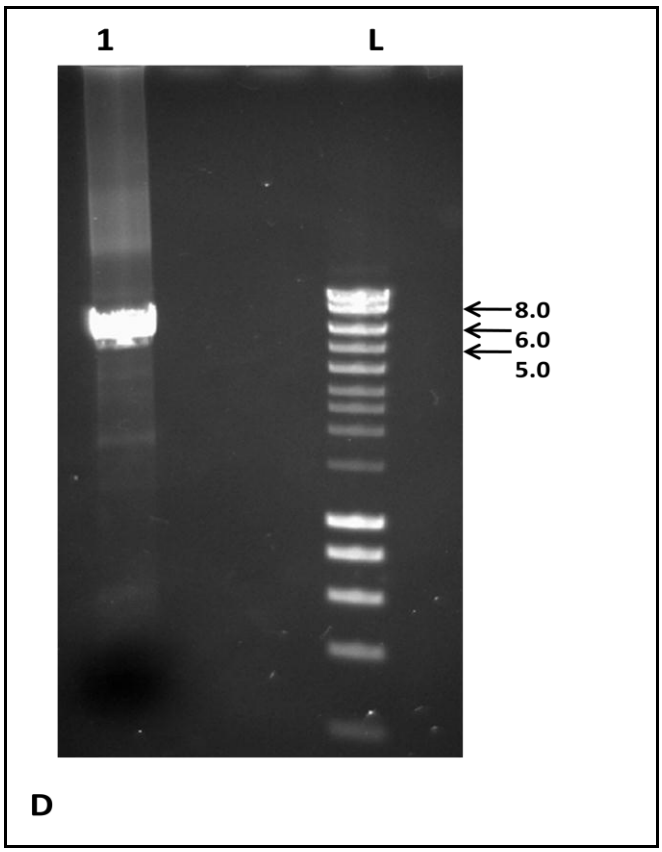


Fig. 3.4 D. PCR amplification of LIC vector. Gel image following AGE shows a fragment between 5 and 6 Kb (expected size 5.4 Kb representing the size of pET-28a vector was expected.). The concentration as determined by spectrophotometric quantification was found to be 50 ng/µl.

UGT fragments were cloned such that recognition sites for *Xba*I and *Xho*I flanked the upstream and downstream regions of the target sequences respectively. This was confirmed by restriction endonuclease digestion of isolated pDNA (see section **2.2.1.5** for crude pDNA prep) from a single clone by *Xba*I and *Xho*I (see section **2.2.1.11** for reaction conditions). **Fig. 3.5 A** and **B** show AGE images of r-plasmids harbouring target gene inserts. Internal restriction recognition sites were present in gene fragments of Pg (*Xba*I and *Xho*I recognition sites present at position 759 and 1136 respectively), Pp (*Xho*I recognition site present at position 68) and Dh (*Xba*I recognition site present at position 717) as determined by the NEBcutter program (http://tools.neb.com/NEBcutter2/). As shown from Lanes 7-9, due to the presence of an *Xba* I recognition site in the case of Dh, two fragments are observed approx. 900 bp and over 1000 bp. Similarly, in the case of Pg three fragments at approx. 800, 600 and 400 bp are observed as expected based on the positions of recognition sites. In all other cases, bands of the expected gene size were observed with the addition of approx. 100 nucleotides which includes LIC specific ends as well as addition of nucleotides until the restriction site for *Xba* I is included in the insert. Linearised LIC vector appeared as a band of approx. 5 Kb in all cases. Results of successful cloning of target inserts are tabulated in **Table 3.8**. These results clearly indicate that target inserts from all sources were successfully cloned in the LIC vector.

Organism	Total pDNA preps screened for inserts	No. of r-plasmids	Recombination efficiency (%)
Kp	7	1	14
Pa	5	4	80
Sc	4	3	75
Dh	4	3	75
Kp	5	5	100
Pg	8	7	88
Pp	6	6	100

Table 3.8 Recombination efficiency of plasmids. Constructs were transformed in *E. coli* (TOP10) and efficiency verified by restriction endonuclease digestion.

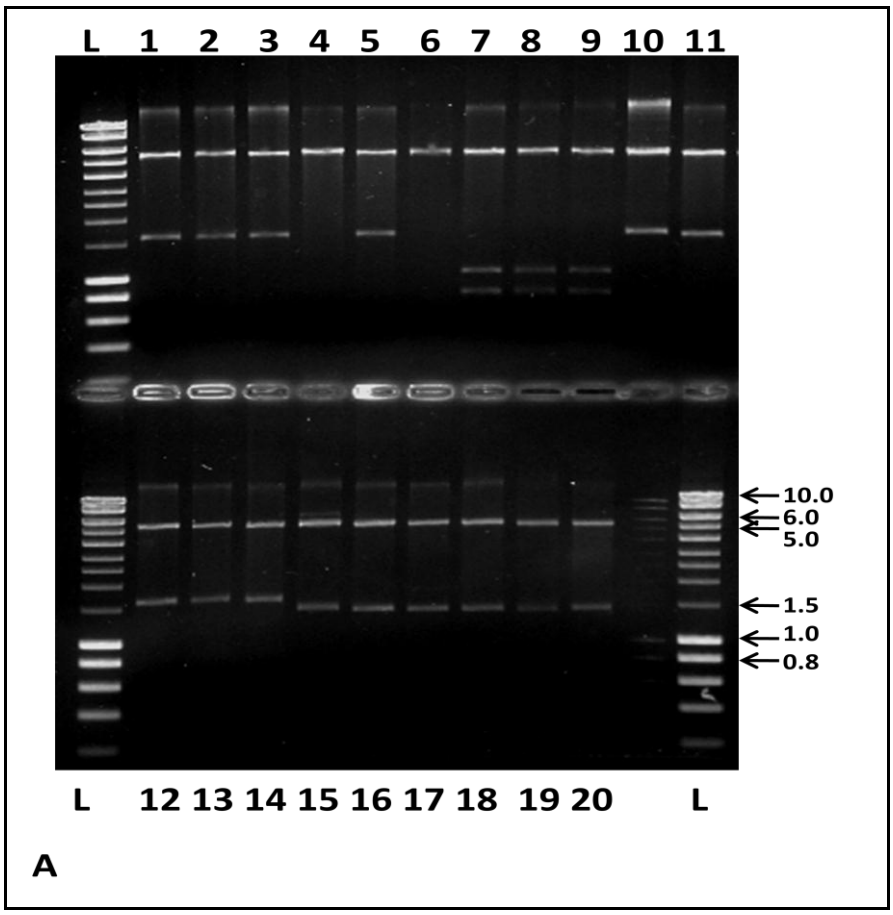


Fig. 3.5 A. AGE gel image showing gene fragments following endonuclease digestion of pDNA. *Lanes 1-5:* Pa gene fragments (expected size 1513 bp); *Lanes 6-9:*Dh gene fragments; *Lane 10:* Kl gene fragments (expected size 1573 bp)*; Lane 11:* Kp gene fragment (expected size 1534 bp); *Lane 12-14:* Sc gene fragments (expected size 1531 bp); *Lanes 15-20:* Pp gene fragments (expected size 1495 bp); *Lanes 4, 6:* pDNA lacking target gene fragments. L= DNA size standards ranging between 200 bp and 10 Kb.

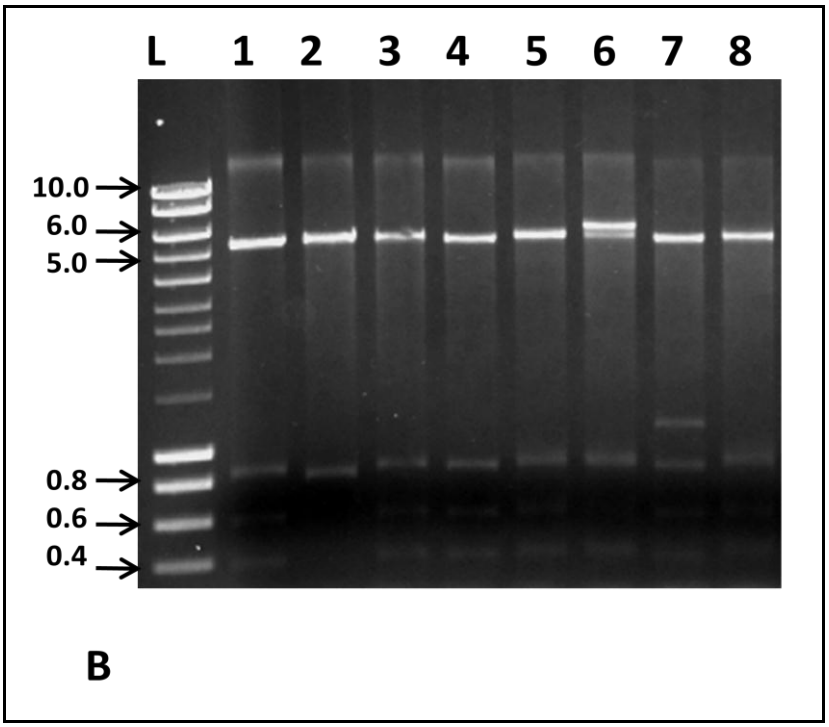


Fig. 3.5 B. AGE gel image showing fragments following endonuclease digestion of pDNA. *Lanes 1-8:* Pg digested DNA showing multiple fragments due to presence of internal restriction endonuclease sites.

3.3.2.2 Native target gene expression and protein purification

A single r-plasmid of each gene was transformed into chemically competent *E. coli* (DE3) for protein expression. All clones were transformed in *E. coli* (DE3) strains BL21, C41, C43, B834, CP (CP) and CP (RIPL) in order to optimise protein expression. Additionally, in order to determine optimum growth temperature for maximal protein expression, *E. coli* cultures were grown at 23^{o}C, 30^{o}C and 37^{o}C post induction with IPTG. This is because change in incubation temperature can significantly alter protein expression (Yu *et al.* 2006). The optimum temperature for protein expression in all cases was found to be 23^{o}C.

Protein expression profiles were compared to determine the most appropriate *E. coli* (DE3) strain for maximal expression. **Fig. 3.6** shows SDS-PAGE analysis of solubilised proteins from overnight cultures of DE3 strains harbouring the pPpUGT51'. The PpUGT51' (m.wt 54 kDa) was expressed, with the highest level of expression observed from CP (RP, RIPL) strains. *E. coli* CP (DE3) strains are engineered to contain additional t-RNA encoding genes to increase heterologous protein expression in *E. coli* as has been shown previously (Singh *et al.* 2009). Small scale protein expression levels in LB and auto-induction media (section **2.2.1.3**) were compared to check if media could affect protein hyper-expression. It was found all r-proteins expressed as insoluble inclusion bodies using auto-induction media as compared with LB media which proved better for soluble protein expression (SDS-PAGE gel image using auto-induction media in **Appendix C.**). Furthermore, membrane fractions and CFEs of samples harvested 3 h post induction with IPTG in addition to overnight incubation post induction

were analysed and in all cases it was found that soluble protein expression increased post overnight incubation following IPTG induction.

Negative controls were performed wherein no IPTG was added to induce protein expression. In these cases no protein over-expression was observed as compared to cultures induced with IPTG. **Table 3.9** indicates the most appropriate *E. coli* strain for maximal protein expression for each construct (See **Appendix E.** for SDS-PAGE gel images for r-proteins of Sc, Pa, Kl, Kp, Pg and Dh preferentially over expressed in indicated strains). In each case, optimised strains were then used for consequent protein expression and purification.

CFEs (section **2.2.2.2**) were prepared from IPTG induced overnight *E. coli* cultures harbouring r-plasmids. *N*-terminally histidine tagged r-proteins were purified from CFEs using gradient elution IMAC with a Ni Sepharose high performance resin. An automated gradient elution technique (as described in section **2.2.2.3**) was used wherein relatively low imidazole concentration (10 mM) was used to eliminate *E. coli* native proteins while the *N*-terminal hexa-histidine tag engineered on the r-proteins chelates to the Ni^{+2} ions and is eventually eluted by introducing a linear gradient of 10-500 mM imidazole (Yang *et al.* 2007). This automated purification method proved highly efficient in protein purification. Even though KpUGT51' was solubly expressed, it could not be purified successfully via IMAC. All other r-proteins were purified to near homogeneity. Pooled purified fractions were pooled, visualised by SDS-PAGE as shown in **Fig. 3.7**. These fractions were concentrated and protein yields were between 4-6 mg/L. All enzymes were stable for up to 4 months at $4^{o}C$.

r-protein	Optimised *E. coli* strain/strains for soluble protein expression
PpUGT51'	CP (RP, RIPL)
PgUGT51'	CP (RP, RIPL), B834
PaUGT51'	CP (RP, RIPL), B834, BL21
ScUGT51'	CP (RP, RIPL)
KlUGT51'	CP (RP, RIPL), BL21
KpUGT51'	B834, BL21
DhUGT51'	CP (RP, RIPL), BL21

Table 3.9 *E. coli* (DE3) strains for soluble protein hyper expression. *E. coli* strains showing the maximal soluble protein expression levels for each r-plasmid. It is worth noting that no expression was detected in the conventional *E .coli* (BL21) for PpUGT51', PgUGT51' and ScUGT51'.

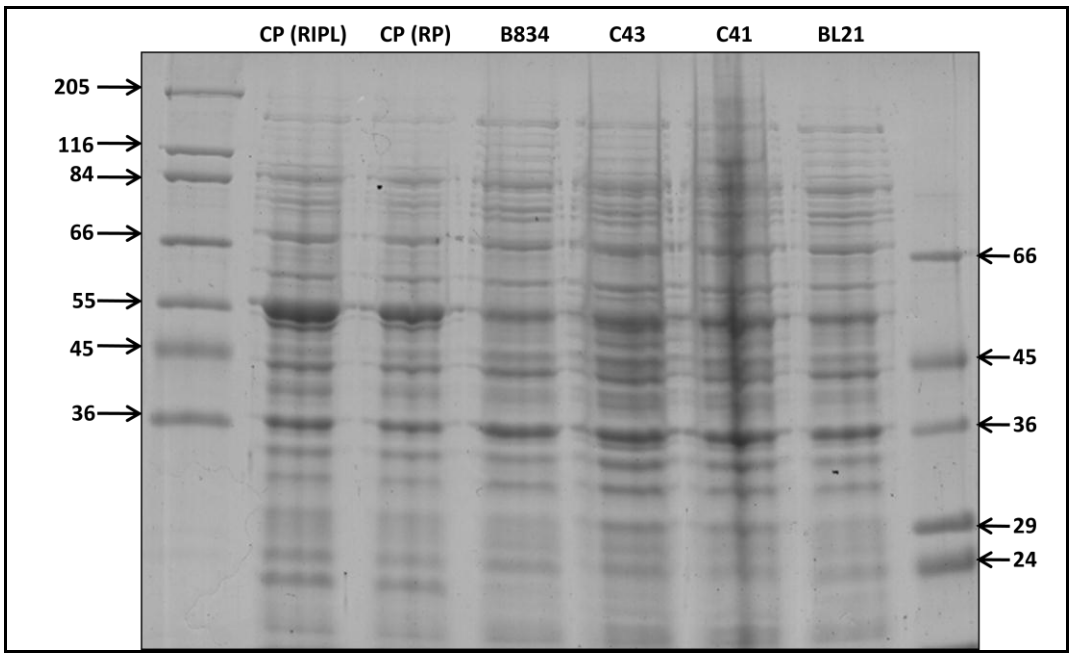


Fig. 3.6 Optimisation of protein expression levels in *E. coli* (DE3) strains. SDS-PAGE analysis of membrane fractions of *E. coli* (DE3) (indicated above lanes) following expression of the Pp r-plasmid. Higher expression levels can be seen in case of *E. coli* CP (RP, RIPL). Predicted m.wt for PpUGT51'-54 KDa.

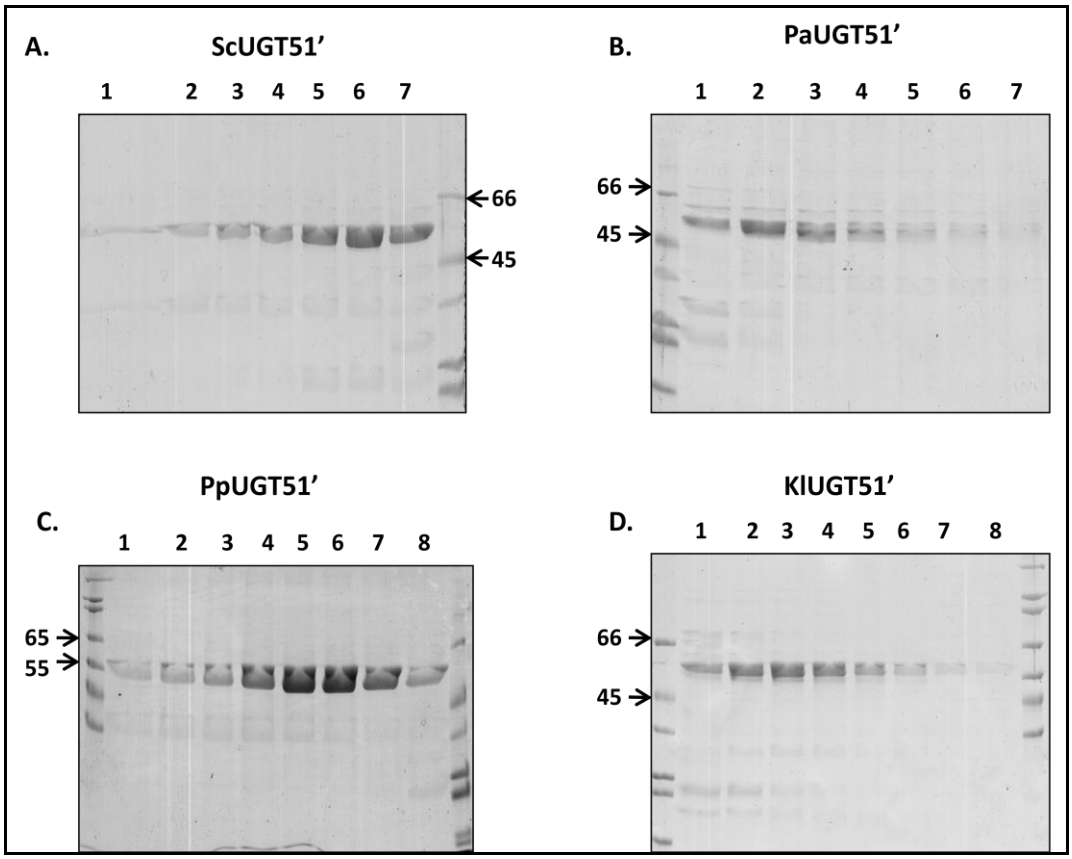


Fig. 3.7 SDS-PAGE analysis of IMAC purified UGTs. Elution of UGTs was achieved at imidazole gradient ranging between 10 and 500 mM. Purified protein bands are observed for **A.** ScUGT51' (predicted m.wt-55 KDa); **B.** PaUGT51' (predicted m.wt- 54 KDa); **C.** PpUGT51' (predicted m.wt 54 KDa) and **D.** KlUGT51' (predicted m.wt 54 KDa). Gel images for purified fractions of Pg and Dh r-proteins are included in **Appendix E**.

3.3.3 Determination of *in vitro* UGT activity and kinetic analysis

3.3.3.1 In vitro UGT activity assays and LC-MS analysis

To determine the preliminary activity and substrate specificity of UGTs, *in vitro* assays (section **3.2.5**) were performed with the seven solubly expressed proteins, six of which could be successfully purified as described in section **3.3.3.** For ease of reference, r-proteins produced using truncated sequences of identified yeasts will be designated as PpUGT51', PaUGT51', KlUGT51', PgUGT51', KpUGT51' and DhUGT51' isolated from Pp, Pa, Pg, Kp and Dh respectively. The respective r-plasmids will be denoted with a prefix 'p'; for example, pScUGT51' for the r-plasmid of ScUGT51'.

Protein concentrations were determined by measuring the A_{280} values as described in section **2.2.2.5.** Based on previous studies of GTs isolated from eukaryotes, *in vitro* UGT activity assays were performed at pH 8.0 and a temperature of 30°C (Madina *et al.* 2007; Warnecke *et al.* 1999). No activity was observed when assays were performed at 37°C although blood group GTs have been previously assayed at this temperature (Persson and Palcic 2008). To test UGT activity and substrate specificity, assays were performed with a panel comprising mainly of steroidal acceptor substrates and UDPG and UDPGA as glycosyl donors. Structures of acceptors tested for UGT activity are depicted in **Table 3.10 A.** The range of acceptors served as examples with varying degree of steric hindrance apart from the difference in positioning of the –OH functional group for glycosidic bond formation. These were ergosterol, cholesterol, dihydroxycholesterol (cholestanol) and testosterone. β-naphthol, a non-steroidal acceptor substrate with a –OH functional group was also included to assess the

promiscuity of UGTs. It differed from steroidal substrates in that it consists of two fused benzene rings as opposed to four fused cycloalkane rings, a characteristic feature of steroids.

Preliminary UGT activity results are summarised in **Table 3.10 B**. Assays (section **3.2.5**) were set up in triplicate and incubated for 2 h following which the lipids were extracted, dried to evaporation and reconstituted in 2:1 (v/v) chloroform: methanol. The extracted lipids were then observed on a TLC plate following heat charring. KpUGT51' could not be purified by IMAC and UGT activity assays were performed with 60 µl CFE in this case. As expected, ScUGT51' displayed USGT activity. In addition, three r-proteins encoded by the selected gene fragments of Pp, Pa and Kl also displayed USGT activity with testosterone and/or ergosterol, cholesterol and cholesterol derivatives. No product was observed when assays were performed using non-steroidal aglycones such as beta naphthol. Furthermore, UDPG was shown to be the universal donor substrate that could not be replaced by UDPGA. UGT activities were further verified by assays with CFEs using UDP*G* as the donor and cholesterol as the acceptor aglycone (section **2.2.3**). Untransformed *E. coli* BL21 (DE3) CFEs and crude membrane fractions were used as a control and it was found that these did not show background SG activity along with r-proteins from Pg, Dh and Kp with dpm counts typically less than 100. As expected, Sc, Pp, Pa and Kl UGTs were active with dpm counts between 1000-2000.

The reaction products with cholesterol, testosterone and ergosterol, resolved by TLC in a 85:15 (v/v) chloroform: methanol solvent mixture are shown in **Fig. 3.8**. The reaction products with testosterone as substrate were co-chromatographed

with a chemically synthesised testosterone glycoside (TG) standard (ProStrakan) and were found to have identical Rf values. The standard and sample also had identical Rf values when co-spotted and visualised by TLC. Control reactions lacking UDPG showed the presence of the unused acceptor while lacking the formation of any other reaction products.

Following in-gel trypsin digestion and MS analysis, the identity of the active r-proteins could be verified by a Mascot nr peptide database search (described in section **3.2.4)**. The MS data subjected to searches against the nr protein sequence database confirmed the successful identification of Sc, Pp, Pa r-proteins based on the detection of 10, 14 and 10 peptide sequences respectively. The Mascot spectrum analysis reports are included in Appendix **B**.

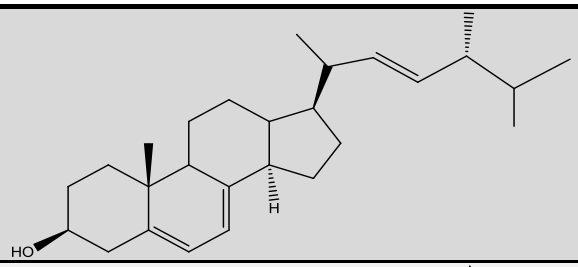
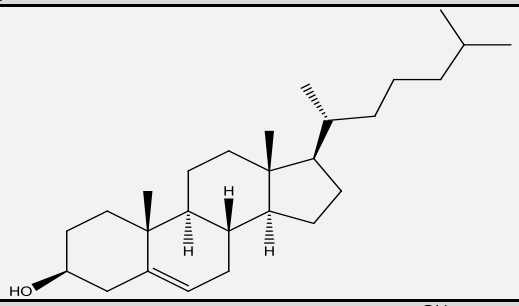
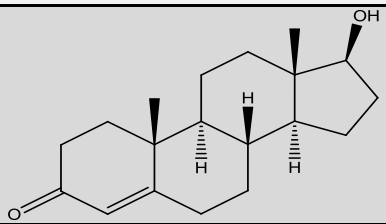
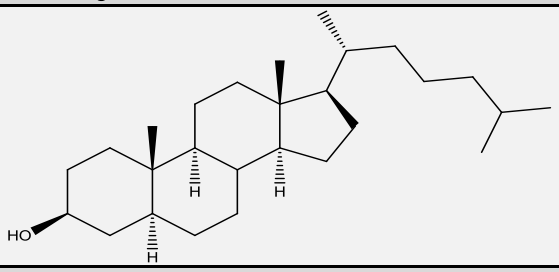
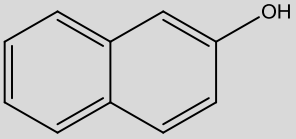
Code	Acceptor substrate	Structure
E	ergosterol	
C	cholesterol	
T	testosterone	
DC	5-alpha-cholestan-3-beta-ol (dihydrocholesterol)	
B	β-naphthol	

Table 3.10 A. Structures of acceptor substrates used to study UGT activity.

r-protein	Acceptor substrates	Glycosyl donor
ScUGT51’	E,C, DC, T	UDPG
PpUGT51’	E,C, DC, T	UDPG
PaUGT51’	E,C, DC, T	UDPG
KlUGT51’	C, DC	UDPG
KpUGT51’	-	-
PgUGT51’	-	-
DhUGT51’	-	-

Table 3.10 B. Summary of UGT activity against tested substrates. Acceptors utilized by UGTs as confirmed by TLC analysis are mentioned. Codes designated to acceptors are mentioned in **Table 3.10 A**.

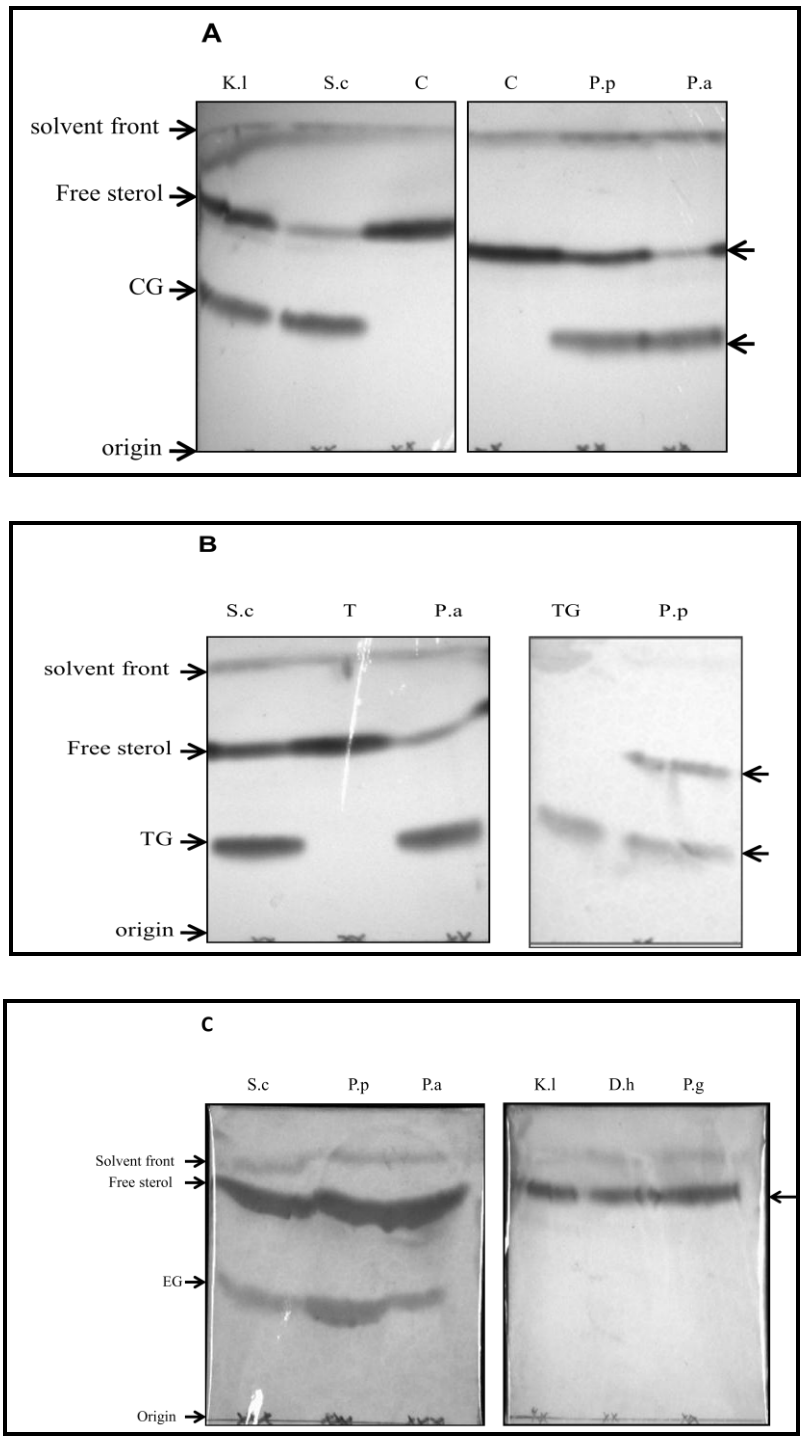


Fig. 3.8 TLC resolution of extracted reaction products. Products were analysed following 2 h incubation of Sc, Pp, Pa and Kl UGT51' reaction mixtures with **A.** cholesterol and **B.** testosterone and **C**. ergosterol acceptors. Lanes C and T are controls without UDPG. Lane TG depicts the authentic standard β-testosterone glycoside. Unused acceptor substrate and respective SGs are indicated by upper and bottom arrows and were detected by heat charring following staining with 15% (w/v) phosphotungstic acid.

For LC-MS analysis and verification of the TLC-resolved reaction products with testosterone as the acceptor (section **2.2.4)**, reaction volumes of the above described assays were doubled and after stopping the reaction after 2-3 h, half the volume of the reaction mixture, i.e. 100 µl was directly loaded for LC-MS whereas the other half was processed for TLC analysis. The expected *m/z* ratio for testosterone glycoside was 451.0 and 289.0 for testosterone (in the positive mode). The analysis was performed in the SIM mode (288.0-290.0; 450.0-452.0) for these ions and the chromatograms with the mass spectra for each UGT are shown in **Figs. 3.9 (A.-C.)**. Retention times of reaction products were compared using standards of testosterone and testosterone glycoside (TG) which showed similar retention times of 2.0 and 1.3 min respectively. Also, control assay reactions with UDPGA as glycosyl donor or reactions in the absence of either substrate (acceptor or donor) showed no formation of a product peak. These results therefore confirmed TG synthesis by the action of UGTs of Sc, Pp and Pa. Following preliminary TLC analysis where no TG was noted using KlUGT51', LC-MS analysis also did not yield a TG product peak following KlUGT51' assays. Although LC-MS has been routinely used to detect testosterone and its metabolites, elongated run times have been reported (Li *et al.* 2002; Zea and Pohl 2004). In this study, detection of testosterone and the corresponding conjugated product was achieved within a total run time of 5 min in isocratic mode thereby making this is rapid screening system for active GTs which may also be amenable to HTS.

Glycosyl transfer reactions proceeded to completion as no characteristic peak was detected at *m/z* 289.0 following overnight incubation of reaction mixtures, thereby

indicating complete conversion of testosterone. This was additionally confirmed by the TLC resolution of the extracted reaction products where no remaining acceptor substrate was detected. The MS monitored increase in TG peak area using reactions stopped at different time intervals could therefore possibly be used for kinetic analysis of UGTs against a standard curve produced using known amounts of TG. A few studies have reported determination of GT kinetic parameters by LC-MS (Bothner *et al.* 2000; Cartwright *et al*. 2008; Hefner and Joachim 2003) and LC-MS (Offen *et al.* 2006). In one such study by Cartwright *et al*. (2008), an HPLC based assay was reported to determine kinetic parameters of flavonoid specific UGTs from *A. thaliana*. Each of these UGTs exhibit differential preferences towards the different –OH groups in quercetin and the differences in regiospecificity were exploited to determine kinetic parameters. However, the mentioned studies report significantly lower K_{cat} (< 2.9 s^{-1}) and catalytic efficiency compared to USGTs in this study. This could be due to the high sensitivity of radiochemical assays over conventional methods for analysis (Brazier-Hicks *et al.* 2007). Also, in the absence of an internal standard, the requirement for manipulation for easy detection such as derivitization of either substrate coupled with lack of reproducibility make the accuracy of MS based quantification studies debatable (Higashi and Shimada 2004). Yang *et al*. (2005) have described the quantification of product accumulation over time catalysed by GT family 1 plant UGTs by using GDP as an internal standard. However, NDPs have been known to cause GT inhibition and this may limit their further use as internal standards for monitoring reactions in such cases (Potocka and Zimowski 2008). LC-MS analysis could only be performed using testosterone as the acceptor as cholesterol represented a poor candidate for MS analysis due to its

ionization response. Even though gas chromatography coupled to mass spectrometry (GC-MS) has been earlier used to detect free sterols (Sugai *et al.* 2009), monitoring the corresponding glycoside formation still represents a major challenge in analyzing GT catalysed reactions.

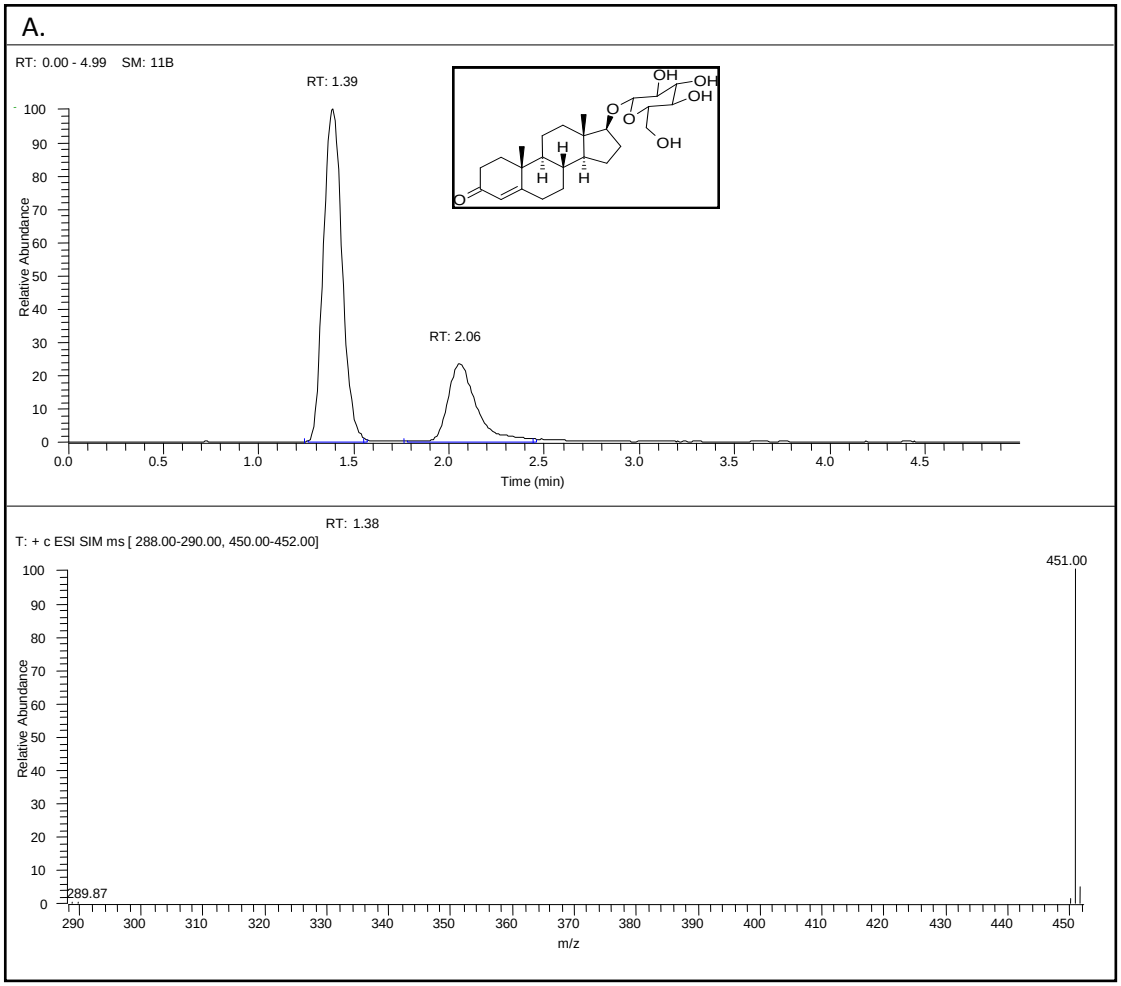


Fig. 3.9 A. LC-MS analysis of PaUGT51'- testosterone assay products. Products were analysed following a 2 h incubation of reaction mixture. A characteristic peak at m/z = 451 $[M + H]^+$ at a retention time (R.T.) of 1.3 min confirms the synthesis of TG (inset); peak at R.T. 2.06 indicates unused testosterone. Ion with m/z=289 indicates the breakdown product ion of TG in the reaction.

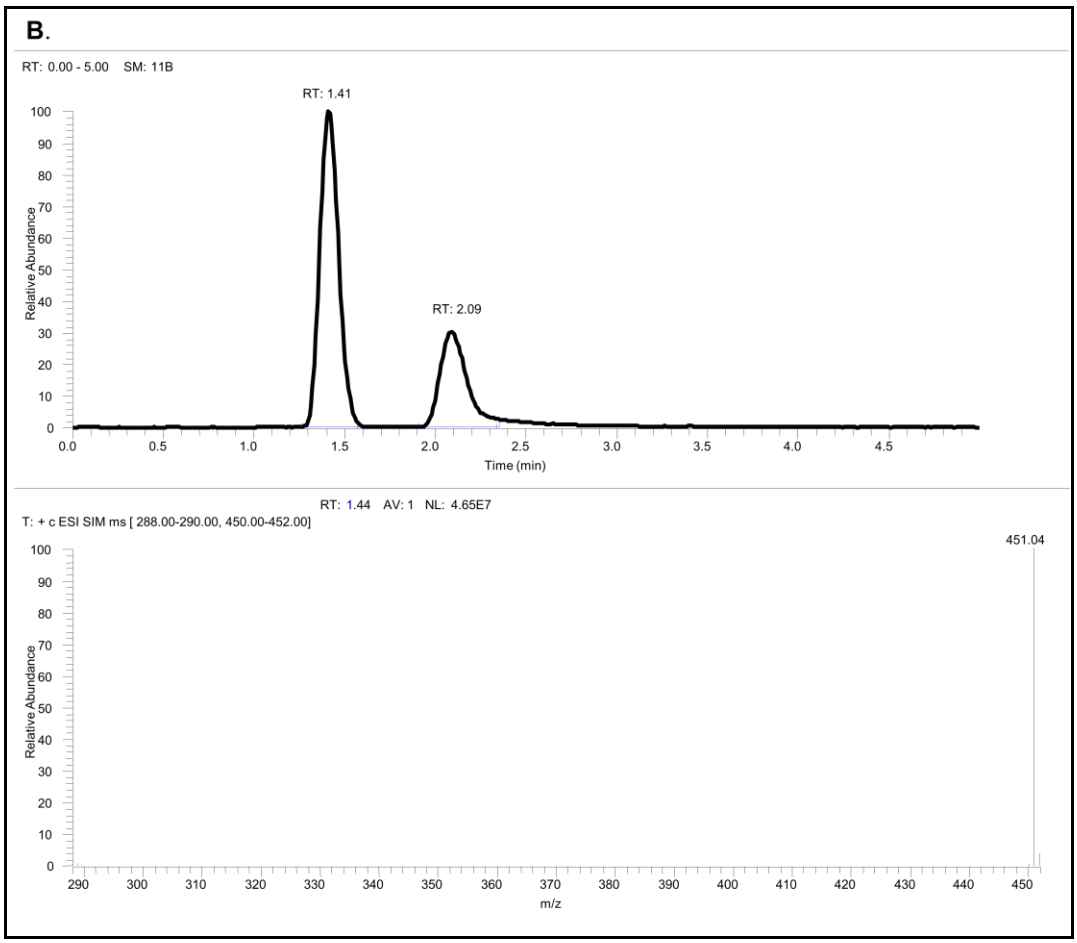


Fig. 3.9 B. LC-MS analysis of ScUGT51'- testosterone assay products. Products were analysed following a 2 h incubation of reaction mixture. A characteristic peak at m/z = 451 $[M + H]^+$ at a retention time (R.T.) of 1.4 min confirms the synthesis of TG (inset); peak at R.T. 2.06 indicates unused testosterone. Chromatogram below shows the TG ion.

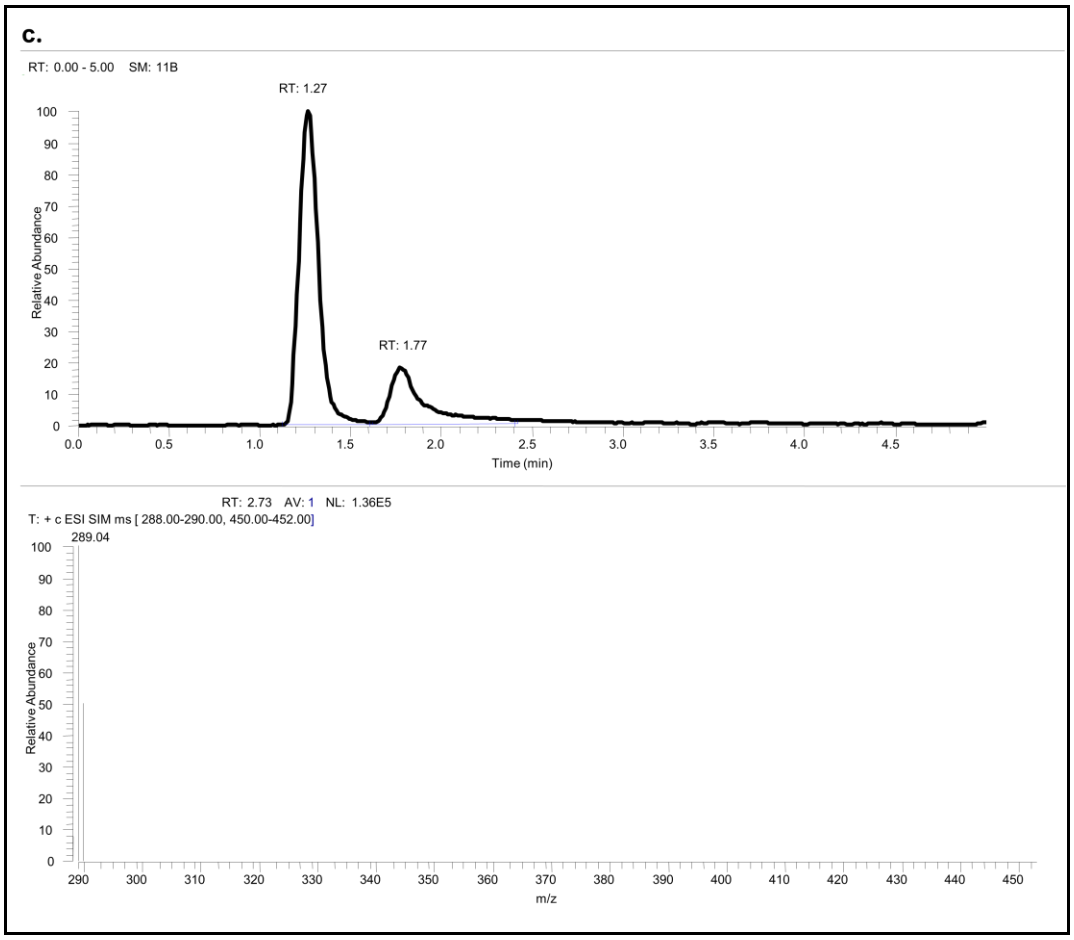


Fig. 3.9 C. LC-MS analysis of PpUGT51'- testosterone assay products. Products were analysed following a 2 h incubation of reaction mixture. A characteristic peak at m/z = 451 $[M + H]^+$ at a retention time (R.T.) of 1.27 min confirms the synthesis of TG (inset); peak at R.T. 1.77 indicates unused testosterone. Ion with m/z=289 indicates the breakdown product ion of TG in the reaction.

3.3.3.2 Kinetic analysis of UGTs with cholesterol and testosterone

Assays for kinetic analysis were prepared using pure UGTs as described in section **2.2.3**. Kinetic parameters (K_M, k_{cat}) and catalytic efficiency of all active UGTs were determined as described in section **3.2.7** against a sterol (cholesterol) and steroidal (testosterone) acceptor with UDP*G* as the sugar donor. Kinetic analysis was performed using GraphPad Prism version 5.04 for Windows (GraphPad Software, San Diego California USA) and the results are tabulated in **Table 3.11**. All reactions followed typical Michaelis-Menten kinetics and the addition of $MnCl_2$ in the reaction buffer did not alter enzyme activity. The Lineweaver-Burk plots for all UGTs against cholesterol and testosterone are shown in **Figs. 3.10** and **3.11** respectively. In all cases, under the current experimental conditions, cholesterol was at least two-fold more readily accepted compared to testosterone and this was further highlighted by at least four fold lower K_M (high affinity) values for cholesterol. However, due to the lack of solubility of testosterone at higher concentrations in the aqueous reaction mixture, the K_M values using testosterone ('pseudo' K_M) were derived at less than saturating substrate concentrations. This high K_M value therefore also suggests substrate inhibition. All USGTs exhibit preferential specificity towards the 3-hydroxy group in the A ring of the planar 4-ring system of cholesterol. PpUGT displayed the highest level of catalytic efficiency against both substrates with values of 111.7 s^{-1} μM^{-1} against cholesterol and 2.31 s^{-1} μM^{-1} against testosterone.

UGT	Acceptor	K_M (µM)	V_{max} (µmol $min^{-1}mg^{-1}$)	k_{cat} (s^{-1})	k_{cat}/K_M (s^{-1} $µM^{-1}$)
Sc	C	79.9 ± 21.7	1447.0 ± 218.2	1319.2	16.5
	T	362.5 ± 26.85	538.7 ± 23.75	491.13	1.35
Pp	C	34.90 ± 10.80	4337 ± 551.2	3897.0	111.6
	T	211.1 ± 27.5	288.7 ± 20.5	488.67	2.3
Pa	C	67.2 ± 31.63	719.1 ± 176.8	650.6	9.7
	T	837.9 ± 182.23	540.1 ± 76.24	351.9	0.42
Kl	C	65.92 ± 31.09	509.0 ± 124.6	475.0	7.21

Table 3.11 Kinetic parameters of wt UGTs. Kinetic constants derived from assays using UDP*G*. Results are the means of triplicated determinations ± standard error (SE).

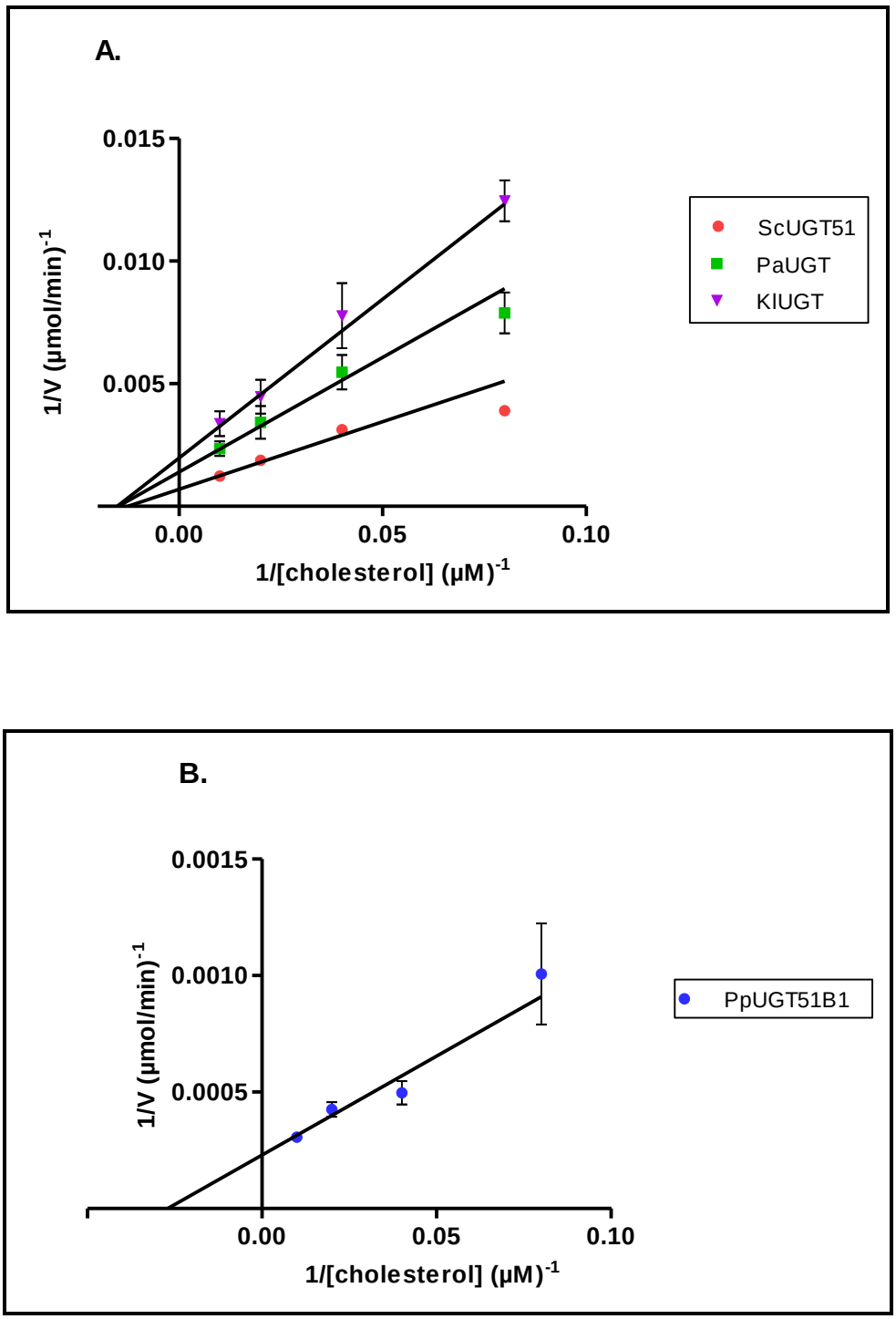


Fig. 3.10 Lineweaver-Burk plots against cholesterol. A. ScUGT51', PaUGT51' and KlUGT51'; **B.** PpUGT51'. Initial reaction rates were determined at [cholesterol] =12.5-100 µM. Data points represent the mean of triplicates; error bars represent standard error of the mean (SEM) of triplicates. V= Rate/mg enzyme.

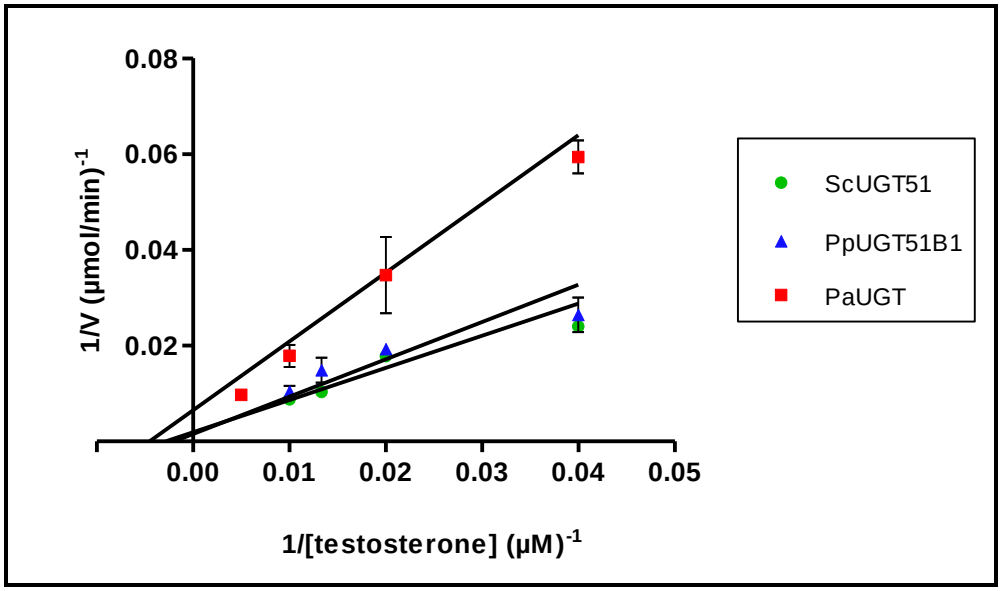


Fig. 3.11 Lineweaver-Burk plots against testosterone. Initial reaction rates were determined at [testosterone] =25-100 µM for ScUGT51' and PpUGT51' and at [testosterone] = 25-200 µM for PaUGT51'. Data points represent the mean of triplicates; error bars represent standard error of the mean (SEM) of triplicates. V= Rate/mg enzyme.

3.4 Discussion

There is limited knowledge on detailed structural information as well as a lack of genetic approaches to elucidate functions of UGTs. ScUGT51' and PpUGT51' are classified as GT family 1 members and therefore can be assumed as inverting GTs that are proposed to adopt the GT-B fold in their tertiary structure. GT members from this family are known to embrace a diverse range of substrates and consist of the characteristic *C*-terminus USS that is involved in NDP-sugar binding (Hansen e*t al.* 2009). For these reasons, characterised members of this family may be used to search sequence databases to identify novel and more promiscuous GTs as well as to understand catalytic domains for this important group of enzymes.

Within the CAZy classification system, the same GT family may display variability in TM content (Rosen *et al.* 2004). Most GT family 1 UGTs characterised so far lack the presence of membrane spanning domains. They do however, contain several hydrophobic segments to enable association with the inner membrane, with the catalytic domain exposed to the cytoplasm (Brockhausen *et al.* 2008). PH and GRAM domains of UGTs from several yeasts such as *P. pastoris* have been shown to be involved in GT localisation to plasma and vacuolar membranes (Oku *et al.* 2003). In addition, the presence of these domains is vital for cellular processes such as pexophagy in yeasts such as *P. pastoris* and *Y. lipolytica* (Stasyk *et al.* 2003).

The *N*-terminus PH and GRAM domains of ScUGT51' are membrane localising elements and therefore are located in low levels in the soluble fractions. Grille *et al*., (2009) have recently reviewed that the majority of UGTs characterised so far

have been based on assays with crude cell homogenates or partially purified preparations which hampers their detailed biochemical and structural analysis.

This study focussed on ScUGT51', previously recognized as a USGT with known function (Warnecke *et al.* 1999) and the identified related UGTs. The full length ORFs of all sequences using ScUGT51' as the query sequence consisted of the PH and GRAM domains; however, the 5' truncated ORFs that did not encode these elements expressed soluble *N*-terminally histidine tagged soluble proteins that were amenable to purification and subsequent analysis. Based on this, apart from the previously characterised ScUGT51' and PpUGT51', the activity of two novel UGTs namely the PaUGT51' and KlUGT51' was described in this study. In all cases where activity was shown, it has confirmed that the majority of the *N*-terminal domain is not required for UGT activity. Although the Atg protein homologs of Dh and Pg (DhUGT51' and PgUGT51' respectively in this study) have been implicated in autophagy and related processes (Bellu and Kiel 2003; Weijers *et al.* 2008), they do not demonstrate UGT activity as shown in this study. However, in the case of Pg, Kp and Dh r-proteins that lacked activity against the tested substrates despite significant sequence similarity with ScUGT51', further structural information may be required to highlight the role of specific residues involved in acceptor specificity.

Interestingly, Sugai *et al.* (2009) have recently reported the detection of an increased amount of total SG content detected in *K. lactis* strain M-16 as compared to strain NBRC 1267. The amino acid sequence of the latter is identical the sequence of the UGT of NRRL Y-1140 used in this study although they both differ from the UGT of M-16 in only one residue at position 620 which is

threonine (T) whereas the residue is methionine (M) in the case of M-16. Even though methionine has a more hydrophobic side chain as compared to threonine, the increased levels of SG in the case of M-16 is not clear. Moreover, the KlUGT51' studied here was void of the *N*-terminus threonine residue and still displayed a relatively lower catalytic efficiency as compared to other UGTs and so the role of this residue in the regulation of SG cannot be confirmed. In addition, the possibility of M-16 being more resistant to conditions has been suggested as such conditions trigger activation of USGTs (Sugai *et al*. 2009). On the other hand, due to the identical primary UGT sequence of Kl strain NRRL Y-1140 and NBRC 1267, other sterols involved in the yeast biosynthetic pathway such as lanosterol, zymosterol and episterol but not included in this study, may also be implied for SG synthesis in NRRL Y-1140 as shown in the case of strain NBRC 1267.

Substrate specificity of UGTs and the effect of divalent cations

In order to evaluate UGT specificity, in addition to the natural acceptor ergosterol, cholesterol and a panel of other such steroidal compounds were chosen as models for xenobiotic sterols in the current study. Sugar linked sterols have tremendous potential as drug delivery tools and as anticancer therapeutics. For example, Hahismoto *et al*. (2008), have successfully reported the inhibitory activity of mono-, di and tri-saccharide cholestanol compounds against mice tumour cell lines (Hahismoto *et al.* 2008).

All active UGTs demonstrated exclusive substrate specificity towards their steroidal as well as sugar substrates. Hefner *et al*. (2003) have shown that a plant GT family 1 arbutin synthase involved in the synthesis of natural product *O*-β-D-

glucosides was able to use β-naphthol as the acceptor substrate with UDPG, the most commonly utilized sugar donor in plants. This result is in agreement with the fact that plant natural product UGTs are largely specific to plant species only and display broad substrate specificity towards the acceptor aglycone (Thorsoe *et al.* 2005).

UGTs also show very high specificity towards their donor sugars. The most commonly used natural sugar is UDPG but others recognized include UDPGA, UDP-mannose, UDP-xylose and the commonly used sugar by mammalian UGTs, UDP-glucuronic acid (Osmani *et al.* 2009). GT family 1 glucuronosyltransferases are involved in the biotransformation of various steroid hormones including testosterone (Compain and Martin 2001). Conjugation with glucuronic acid during Phase II metabolism increases the polarity of the aglycone thereby facilitating its removal by excretion (Evans and O'reilly 1998; Jantti *et al*. 2007). Thus, UGTs responsible for testosterone conjugation could provide a basis for steroid based drug design and biotransformation.

As expected, USGTs were highly specific to sterols and showed no activity against the non-steroidal β-naphthol. Previous studies have shown that ScUGT51' and PpUGT51' and other fungal USGTs with UGT51 homologs such as the UGT51E1 from *Leptosphaeria maculans* with a 38% primary sequence identity to PpUGT51' were able to conjugate sterols such as the ergosterol, cholesterol and (Idnurm *et al.* 2003; Oku *et al.* 2003; Warnecke *et al.* 1999). In other cases, plant USGTs have been reported to utilise steroidal alkaloids such as tomatidine and solanidine acceptors (Moehs *et al.* 1997; Potocka and Zimowski 2008). However, all these compounds are hydroxylated at the C-3 position which USGTs are

commonly known to glycosylate and glycosylation at higher positions is rarely reported. In addition to the C-3 sterols, three USGTs in the present study glycosylated testosterone which consists of an –OH group at the C-17 position. Another plant derived USGT involved in secondary product metabolism has been reported to accept testosterone and higher steroids as acceptors (Madina *et al.* 2007). With a K_M value of 170 µM and k_{cat} >5000 s^{-1} against testosterone, it displayed a much higher catalytic efficiency of 29 s^{-1} μM^{-1} as compared to the USGTs in this study where the highest catalytic efficiency of 2.3 s^{-1} μM^{-1} for testosterone conjugation was achieved in the case of PpUGT. On the other hand, it lacked any activity against the C-3 –OH sterols which were readily accepted by USGTs described in this study.

The presence and/or position of double bonds within the steroid molecule as well as the presence and/or length of the acceptor hydrocarbon side chain has also shown to pose an effect on glycosylation rate with longer alkyl side chains acting as better sugar acceptors as compared with steroidal alkaloids (Potocka and Zimowski 2008). Cholesterol with a longer carbon side chain was found to be at least ten times more readily accepted compared to testosterone besides the fact that it consists of the USGT preferred C-3 –OH group.

It is unclear why of the four tested USGTs only KlUGT51' did not show activity towards testosterone (**Fig. 3.8 B**.) despite 64% sequence identity with ScUGT51'. Therefore, sequence similarity approaches alone cannot be used to predict substrate specificities and domain swap or site directed mutagenesis studies would be able to further elucidate the roles of specific amino acids in acceptor recognition.

The presence of Mn^{+2} did not affect UGT activity. This is in agreement with other family 1 UGTs that are not affected by the presence of Mn^{+2} and Mg^{+2}. For example Madina *et al.* (2007) have reported that the presence of divalent metal cations of Mg and Mn did not affect a plant UGT activity. However, UDP was found to be a competitive inhibitor, suggesting that UDP may play an important role in plants. However, there have been several reports of UGT activity being enhanced in the presence of these cations. For example it has been previously shown that Ca^{+2} significantly enhanced a plant USGT activity as compared to Mg^{+2} (Staver *et al*. 1978). In another study, the rate of glucoside formation was enhanced by over two-fold by the addition of Mg^{+2} whereas Ca^{+2} inhibited a promiscuous plant UGT (Hefner *et al.* 2002). Further 3-D structural information and insight into the substrate specificity would be therefore required to better understand the affect of different metal cations on UGT activity.

Role of SGs

Ergosterol is the most abundant sterol in fungi where along with its biosynthetic precursors including lanosterol, zymosterol and episterol, it plays a major role in maintaining membrane integrity and function (Klutts *et al.* 2006; Stasyk *et al.* 2003). However, a majority of the ergosterol exists in the free form, even though it is modified to steryl glycosides in a few yeasts where it may have a specific function. For example, in the case of *P. pastoris,* deletion of UGT51B1 (PpUGT51' in this study) resulted in a complete loss of ergosteryl glucoside synthesis and glucose dependent peroxisome degradation (Oku *et al.* 2003). In this study, all functionally active UGTs, apart from KlUGT51' have shown positive activity towards ergosterol. Sugai *et al.* (2009) have reported ergosteryl

glucoside synthesis in *K. lactis* M-16, even though SGs from the other studied strain NBRC 1267, identical to NRRL Y-1140 in this study, were not characterised although they report abundance of free ergosteryl in NBRC 1267. This study therefore further highlights the diverse roles of steryl glycosides in M-16 and NBRC 1267/NRRL Y-1140 and in other fungi in general.

4. Random mutagenesis and functional analysis of UGTs

4.1 Introduction

As highlighted in the previous chapters, GTs may possess distinct substrate specificities despite a high percentage of sequence similarities. Thus for a broader scope in drug development, GTs, especially members of the GT-B family represent an ideal candidate for DE and engineering platforms due to the diversity of substrates utilised (Love *et al.* 2006). In order to highlight the role of specific residues in substrate binding and monosaccharide transfer, several mutagenesis studies in the past have focussed on creating point mutations based on solved crystal structures (Shao *et al.* 2005; Thorsoe *et al.* 2005). Even though these studies have highlighted the role of specific conserved residues within the *C*-terminal domain, the highly variable *N*-terminal substrate recognising domains remain relatively poorly understood even though novel specificities have been achieved by several domain swapping studies (Hansen *et al.* 2009; Truman *et al.* 2009).

The aim of the studies described in this chapter was to generate random mutant libraries of the isolated ScUGT51', PpUGT51', PaUGT51' and KlUGT51' and to use these as a resource to screen for mutants with significantly improved catalytic activity compared to the parental enzymes. Beneficial mutations within the truncated versions of the USGTs studied here could provide a better insight into the involvement of specific amino acids in substrate recognition and saccharide transfer to the aglycone. Moreover, as previously described, due to low protein expression levels and GTs being membrane bound, it is important to create variants with significantly improved catalytic activities. Furthermore, due to the high conservation of secondary and tertiary structure among the UGTs,

comparative modeling with solved 3-D structures in the form of homology modeling was also explored in this study. Homology modeling represents an attractive tool for biochemical and functional characterisation as well as for rational engineering of GTs for desired properties (Osmani *et al.* 2009).

Herein, mutant libraries of parental UGT encoding gene fragments were generated by ep-PCR using low fidelity DNA polymerases. This was approached in two different ways. One such approach involved ep-PCR of the wild type (WT) ScUGT51', PpUGT51', PaUGT51' and KlUGT51' sequence using Taq DNA Pol along with unequal amounts of dNTPs, whereas in the other approach, ep-PCR of the WT ScUGT51', PpUGT51', PaUGT51' and KlUGT51' sequence was performed using the Genemorph II EZClone Domain Mutagenesis kit (STR), which contains Mutazyme.

The generation of plasmids with expected high rate of mutation frequencies (upto 9-16 mutations/Kb of gene) in the UGT domains involved two rounds of PCRs. The general scheme followed in RM experiments is depicted in **Figure 4.1**. The ep-PCR stage or the mutant megaprimer (MM) synthesis involved the amplification of the target gene from the original plasmid (donor plasmid) by PCR using a low fidelity DNA Pol (Taq or Mutazyme). The second stage or the PCR cloning stage where PCR products, i.e. mutant megaprimers (MM) from the first round were used as primers to anneal to the original donor plasmid was followed by extension using a high fidelity DNA Pol. While using MMT (megaprimers generated using Taq DNA pol), KOD was used for PCR cloning and this was termed as an EZK reaction. While using MMM (megaprimers generated using Mutazyme DNA pol), EZC enzyme mix (Genemorph II EZClone Domain

Mutagenesis kit (STR), was used for PCR cloning and this was termed an EZC reaction.

A HTS system based upon the desired activity of the evolved enzyme represents a major challenge for screening RM created libraries. The action of a GT in the formation of a glycosidic bond is accompanied by a stoichiometric proton release (Yao *et al*. 1998). For example:

$$\text{UDPGA + N-acetylglucosamine} \rightarrow \text{N-acetyllactosamine +UDP+ } \mathbf{H}^{+}$$

It has previously been shown that this pH perturbation can be used as a sensitive colorimetric assay for galactosyltransferases and other GTs using readily available pH indicators such phenol red and bromothymol blue to follow the progress of the reaction (Deng and Chen 2004; Park *et al.* 2009a; Persson and Palcic 2008). It has been postulated that this methodology would be more widely applicable to other GTs and therefore could be amenable to rapid screening of GT activities using a 96-well plate reader.

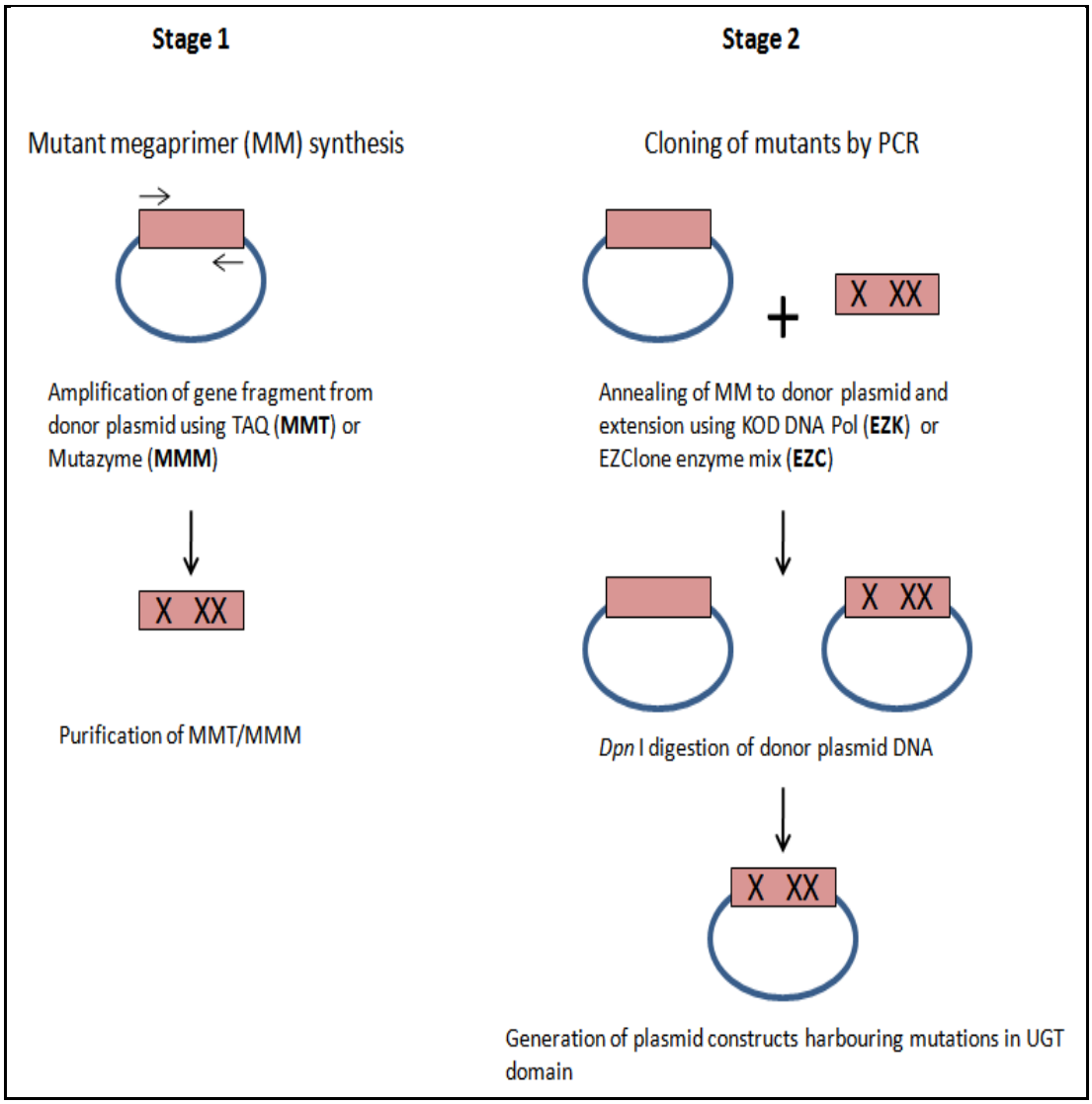


Fig. 4.1 General scheme of RM experiments. Adapted from Genemorph II EZClone Domain Mutagenesis kit instruction manual

4.2 Materials and Methods

4.2.1 Bacterial strains, r-plasmids and sequencing

The r-plasmids pScUGT51, pPpUGT51', pPaUGT51' and pKlUGT51' used are described in Chapter 3. Bacterial growth conditions are described in section **2.2.** DNA was sequenced by GATC Biotech, U.K. For all sequencing reactions, 30 µl DNA prep (50-100 ng/µl) was prepared and samples sequenced using T7 promoter forward (5'-TAATACGACTCACTATAGGG-3') and pET-28a reverse (5'-CTAGTTATTGCTCAGCGG-3') primers.

4.2.2 ep-PCR for MM synthesis using Taq and Mutazyme DNA pol

Primers stated in section **3.2.3.1** were used for ep-PCRs. Stock solutions (100 µM) were diluted to 125 ng/µl (~12 µM). DNA concentration from agarose gels was determined by the Bio-Rad DNA quantification software (Quantity One) against a DNA size standard (Hyper Ladder 1) with known amounts of DNA in each fragment.

4.2.2.1 ep-PCR for the generation of MMT using Taq DNA pol (BL)

Constituents for Taq ep-PCR are listed in **Table 4.1 (a)** and details of the PCR program used is provided in **Table 4.1 (b)**. A template DNA amount of 5 ng was used for high mutation frequency (9-16 mutations/Kb gene fragment).

dNTPs (100 mM each, BL) were diluted prior to use as follows: dATP- 1:10 dilution with H_2O, all other dNTPs- 1:5 dilution with H_2O.

Component	**Volume** (µl)	**Final amount** (in 50 µl reaction)
5.0 U/µl BIOTAQ DNA polymerase	0.5	2.5 U
10x NH_4 buffer [1]	5	1x
wt r-plasmid DNA (template) (10-30 ng/µl)		5 ng
125 ng/µl forward primer	1	125 ng
125 ng/µl reverse primer	1	125 ng
10 mM dATP	1.1	220 µM
20 mM dCTP	2	0.8 mM
20 mM dGTP	2	0.8 mM
20 mM dTTP	2	0.8 mM
50 mM $MgCl_2$ solution	3	3 mM
18.2 MΩ/cm H_2O (to a final of 50 µl)		

Table 4.1 (a) PCR components for ep-PCR using Taq DNA pol. [1] Recipe in **Table 2.6.**

Step	Temp (^{o}C)	Time (h:min:s)	Cycle
1	94.0	00:03:00	Initial denaturation
2	94.0	00:01:00	Denaturation
3	**A**	00:01:00	Annealing
4	72.0	00:01:40	Elongation
5			Step 2-4, repeated 29 times = 30 cycles
6	72.0	00:10:00	Final Extension
7	10.0	Hold	Storage

Table 4.1 (b) PCR program for ep-PCR for random mutagenesis. **A** = optimised annealing temperature as listed in **Table 3.2 (b)**. Elongation time was based on 1min/Kb gene fragment to be amplified.

4.2.2.2 ep-PCR for MMM synthesis using Mutazyme (Genemorph II EZClone Domain Mutagenesis kit)

All reactions and PCRs were set up as described in the kit instruction manual. Constituents for Mutazyme ep-PCR are listed in **Table 4.2 (a)**. A control reaction was also set-up using the positive control plasmid DNA supplied within the kit. 1 µl of the supplied positive control primer mix (125 ng/µl each primer) was used. Also 2 µl of 0.5 ng/µl positive control plasmid DNA was used for a final template amount of 1 ng.

Component	Volume (µl)	Final amount/ concentration (in 50 µl reaction)
2.5 U/µl Mutazyme II DNA polymerase	1	2.5 U
10x Mutazyme II reaction buffer	5	1x
WT r-plasmid DNA (template) (10-30 ng/µl)		5 ng
125 ng/µl forward primer	1	125 ng
125 ng/µl reverse primer	1	125 ng
40 mM dNTP mix	1	200 µM (each)
18.2 MΩ/cm H_2O (to a final of 50 µl)		

Table 4.2 (a) PCR components for ep-PCR using Mutazyme

The PCR program used is detailed in **Table 4.1 (b)** with amendments listed in **Table 4.2 (b)**.

Following ep-PCR, MMs generated were purified as detailed in section **2.2.1.6.**

Step	Temp (^{o}C)	Time (h:min:s)	Cycle
1	95.0	00:02:00	Initial denaturation
2	95.0	00:0:30	Denaturation
3	**A**	00:00:30	Annealing
Follow from step 4 (**Table 4.2**)			

Table 4.2 (b) PCR program for ep-PCR for RM using Mutazyme. **A**= optimised annealing temperature as listed in **Table 3.2 (b)**

4.2.3 PCR cloning of MMs followed by transformation in *E. coli*

Purified MMs generated by ep-PCR (section **4.1.1**) served as primers for the PCR synthesis of plasmids harbouring mutations in the UGT domains.

4.2.3.1 PCR cloning of mutants using MMT as primers- EZK reaction

In this reaction, MMT (section **4.1.1.1**) were annealed to donor plasmids followed by extension using KOD DNA pol. PCR constituents and PCR program are tabulated in **Table 4.3 (a)** and **(b)** respectively.

Component	**Volume** (µl)	**Final amount/ concentration** (in 50 µl reaction)
1.0 U/µl KOD DNA pol	1	1.0 U
10x KOD DNA pol reaction buffer	5	1x
wt r-plasmid DNA (template) (10-30 ng/µl)		50 ng
MMT (50-70 ng/µl)		500 ng
dNTP mix (2 mM each)	5	200 µM (each)
18.2 MΩ/cm H_2O (to a final of 50 µl)		

Table 4.3 (a) EZK reaction constituents

Step	Temp (^{o}C)	Time (h:min:s)	Cycle
1	95.0	00:01:00	Initial denaturation
2	95.0	00:00:50	Denaturation
3	60.0	00:00:50	Annealing
4	68.0	00:13:48	Elongation
5			Step 2-4, repeated 24 times = 25 cycles
7	10.0	Hold	Storage

Table 4.3 (b) PCR program for generation of plasmids harbouring mutant UGTs. Elongation time was based on 2 min/Kb fragment amplification.

4.2.3.2 PCR cloning of mutants using MMM as primers- EZC reaction

In this reaction, MMM (section **4.1.1.2**) were annealed to donor plasmids followed by extension using EZClone enzyme mix (Genemorph II EZClone Domain Mutagenesis kit). PCR constituents for this reaction are tabulated in **Table 4.4.** A control reaction was also prepared using 5 µl of the supplied positive control plasmid DNA template and 250 ng of the positive control MMM generated as described in section **4.1.1.2**. PCR conditions are detailed in **Table 4.3 (b)**.

Component	Volume (µl)	Final amount/ concentration (in 50 µl reaction)
2x EZClone enzyme mix	25	1x
EZClone enzyme solution	3	
wt r-plasmid DNA (template) (10-30 ng/µl)		50 ng
MMM (50-70 ng/µl)		500 ng
18.2 MΩ/cm H_2O (to a final of 50 µl)		

Table 4.4 EZC reaction constituents.

Following PCR cloning, EZC and EZK reaction products were digested with 1 µl *Dpn*I at 37^{o}C for 2 h. Digested reaction products were transformed by electroporation into *E. coli* (TOP 10 and BL21) as described in section **2.1.1.17**. A transformation control was also set up for both sets of reactions with 1 µl of pET-28a (1 ng/µl) (expected to result in approx 4000 CFUs upon transformation). In the case of EZC reactions, an additional control transformation was also performed using the supplied positive control plasmid (Genemorph II Domain Mutagenesis kit) containing the *lacZ* gene. In this case, the reactions were plated onto low-salt LB agar supplemented with Amp, IPTG and X-gal.

4.2.3.3 DNA analysis and sequencing

10 single colonies derived for each clone following RM (section **4.1.2**) and transformation in *E. coli* TOP 10 were picked and cultured for extraction of spin

DNA as described in section **2.1.1.5**. Each DNA prep (2 µl) was digested with restriction endonucleases (section **2.2.1.11**) to confirm the presence of the insert. The samples were then sent for DNA sequencing.

4.2.4 Screening for UGT activity following RM

4.2.4.1 Colorimetric assays for HTS

Colorimetric assays were performed with cresol red (CR), phenol red (PR) and bromothymol blue (BB) as the absorbance change indicators. Stock solutions of colour indicators were prepared in appropriate buffers adjusted to the required pH and were stored at room temperature. Assay conditions for each indicator are listed in **Table 4.5**. All components except the UDP donor were added to a 96-well flat-bottom microtitre plates and mixed. The plates were then pre-incubated for 5 min at 30°C in a Microtek Synergy-HT Microplate reader equipped with an internal temperature control chamber. Following this, the UDP donor was added and absorbance monitored at 1 min intervals for a period of 3 h.

Component	Final concentration (wavelengths used)		
	PR (A_{557})	CR (A_{436})	BB (A_{615})
10 mM Na_2HPO_4 buffer; pH 8.0/ MOPS buffer; pH 7.3	2 mM	1 mM	2 mM
2 mM colour indicator	0.01 mM	0.05 mM	0.1 mM
4 mM acceptor sterol	0.3 mM	0.4 mM	0.8 mM
40 mM UDPG/UDPGal	2 mM	2 mM	2 mM
purified UGT in 5 mM TRIS-HCl pH 8.0	1-2 µg	1-2 µg	1-2 µg
18.2 MΩ/cm H_2O to a final volume of 200 µl			

Table 4.5 pH change colour assays using colour indicators. Assays were maintained at a constant pH 8.0 in the case of CR and PR and pH 7.3 in the case of BB. Control reactions were prepared where the UDPG was substituted with UDPGal. Blanks were prepared with 200 µl 18.2 MΩ/cm H_2O. The wavelength of absorbance for each colour indicator is stated.

4.2.4.2 Screening RM libraries by radioactivity assays with UDPG

For screening, 1 ml LB supplemented with 50 µg/ml kanamycin was added to each well of as many sterile 96-deep-well microtitre plates as required. Individual colonies derived for each clone following RM and transformation in *E. coli* BL21 (DE3) were picked and inoculated into 94 out of 96 wells in each plate. A positive (wt UGT transformed in BL21(DE3)) and negative (pET-28a transformed in

BL21(DE3)) control was included in the remaining two wells of each plate following which the plates were sealed with a breathable self adhesive film (SL).

After cell growth at 37^{o}C for 18 h with shaking at 350 rpm, 100 µl of each culture was transferred to a fresh deep-well plate containing 1 ml LB medium supplemented with 50 µg ml^{-1} kanamycin. The original 'master' plate was then sealed and stored at 4^{o}C. The freshly inoculated plates were incubated at 37^{o}C for 2-3 h with shaking at 350 rpm at which point the OD_{600} was approximately 0.7. IPTG was then added to a final concentration of 0.4 mM followed by the incubation of plates for 18 h at 23^{o}C.

Cells were harvested by centrifugation at 3,000 x *g* for 10 min at 4^{o}C and the pellets re-suspended in chilled 50 mM TRIS-HCl (pH 8.0) containing 10 mg/ml lysozyme. Plates were then frozen at -80^{o}C and thawed at room temperature. Cell debris was collected at the bottom of each well by centrifugation at 3,000 x *g* for 20 min at 4^{o}C. The resulting clear supernatant (CFE) from each well were screened for UGT activity based on assays with UDP*G* and analysis by LSC as described in section **2.2.3.** CFEs with significantly evolved activity towards cholesterol compared to wt UGTs were identified and selected for DNA sequencing. For DNA sequencing, 50 µl of culture from a selected well of the stored 'master' plate was inoculated and cultured in 5 ml LB medium for preparation of spin DNA as described in section **2.2.1.5**. Spin DNA was then transformed into chemically competent *E. coli* TOP 10 (as described in section **2.2.1.12**). An individual colony was picked, inoculated and cultured in 10 ml LB medium for spin DNA preparation for sequencing.

4.3 Results

4.3.1 HTS method development

This study had aimed to exploit the postulated colour change associated with a GT catalysed reaction for HTS. However, limited success was achieved in this methodology following which RM libraries were screened for GT activity by the widely used, reliable and sensitive radiochemical assay. Results from both these approaches are described in the following sections.

4.3.1.1 Colorimetric assay method development

In order to establish a simple and rapid screening system for active UGTs following RM, a pH responsive method based on colour and corresponding absorbance change was employed to test wt UGTs wherein three different colour indicator/buffer combinations (CR/PR and Na_2HPO_4 pH 8.0; BB and MOPS; pH 7.3) were evaluated. However, the sensitivity and reproducibility of this technique relies on a number of factors. Firstly, the buffer strength proves crucial in these assays as weak buffers may increase the sensitivity of the system and lead to pH drops upon enzyme or substrate addition and strong buffers may not yield any colour change due to the increase in buffering capacity of the system (Chapman and Wong 2002). Secondly, the pH range for colour change of the indicator should include the optimum pH range for enzyme activity (Banerjee *et al.* 2003). Also, the number of protons released during the reaction progress may not be sufficient to yield a colour change from red to orange/yellow in the case of PR

and CR and blue to green/yellow in the case of BB. In order to confirm that the buffering capacity of the screening system was appropriate for a colour/absorbance change and to estimate the amount of H^+ required to yield this change, a calibration curve was generated with PR assay components (section **4.2.4.1**) omitting the enzyme and consisting of known amounts of HCl (0-0.2 mM). As shown in **Fig 4.2**, a decrease in A_{557} confirmed that assay components and solvents including ethanol, used for the preparation of acceptor stocks did not interfere with the buffering capacity of the mixture. In addition, a steady colour change from red ([HCl] =0 mM) to yellow ([HCl] =0.2 mM) was also observed in these reactions. However, when assays were performed with pure wt UGTs, no significant absorbance change was noted following 3 h incubation. In addition, control reactions where UDPG was substituted with UDPGal also showed no difference in absorbance. Moreover, no apparent colour difference was noted between reactions and control samples. **Fig. 4.3** shows the BB assays prepared in a 96-well plate format. Negative controls were also prepared with UDPGal as the sugar donor and the r-proteins from Dh and Pg. Again, no apparent colour change was observed in the case of reactions with PgUGT51' and DhUGT51' following 3 h incubation at 30^{o}C. Even though KlUGT51' and PpUGT51' reactions show a slight colour change from blue to green, no differences were observed for ScUGT51' and PaUGT51' reactions. Furthermore, in all cases, no colour differences were observed between reactions and corresponding negative controls. These experiments were repeated and similar results were obtained indicating that the buffering capacity of the system could have been affected upon enzyme addition. On the other hand, CR and PH based assays displayed a better contrast between controls and reactions. This could be due to the fact that CR and PR turn

from red to yellow within a pH range of 8.8-7.0 and 8.0-6.6 (Moris-Varas *et al.* 1999) respectively which better correlates with the optimum pH range for UGT activity as opposed to BB that turns from blue to yellow within pH 7.6-6.0. However, these results were not reproducible further highlighting the fact that the addition of purified enzyme solution may cause sudden pH perturbations. Therefore, in order to avoid precipitation, r-protein solutions were maintained in 5 mM TRIS; pH 8.0 and a 1:200 (v/v) dilution was prepared in 2 mM Na_2HPO_4 buffer; pH 8.0 prior to addition to the assay mixture. However, this further did not translate into a colour change following 3 h incubation at 30^{o}C. These results clearly suggest that a high level of UGT activity is required in order to produce a significant pH shift leading to a colour change. Following assays with pure UGTs without yielding an appreciable colour change, the amenability of this approach for HTS for UGT activity using CFEs could not be validated in my hands.

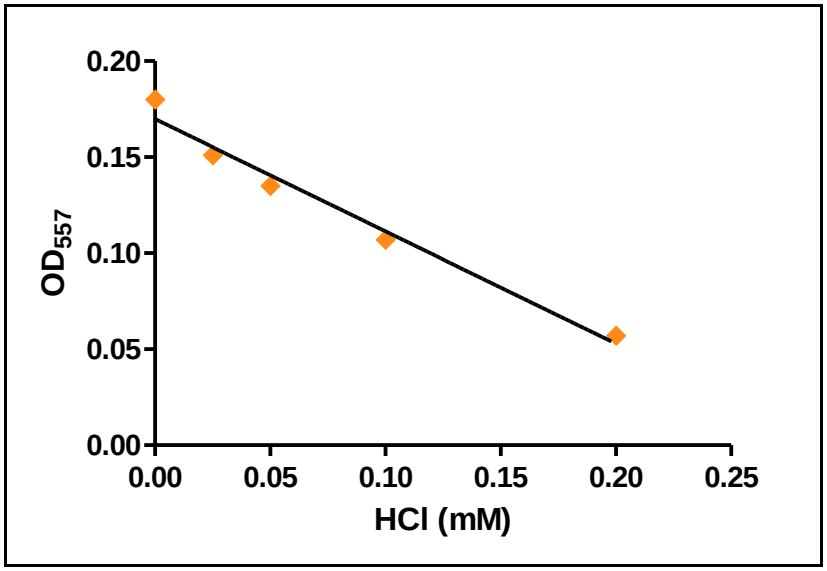


Fig. 4.2 Calibration curve for HCl. Relationship between proton concentration and absorbance at 557 nm. Samples contained (in a total volume of 1 ml): 0.01 mM PR, 2 mM Na_2HPO_4 buffer, 0.3 mM cholesterol and 2 mM UDPG.OD_{557} was recorded after addition of different amounts of HCl (0-0.2 mM).

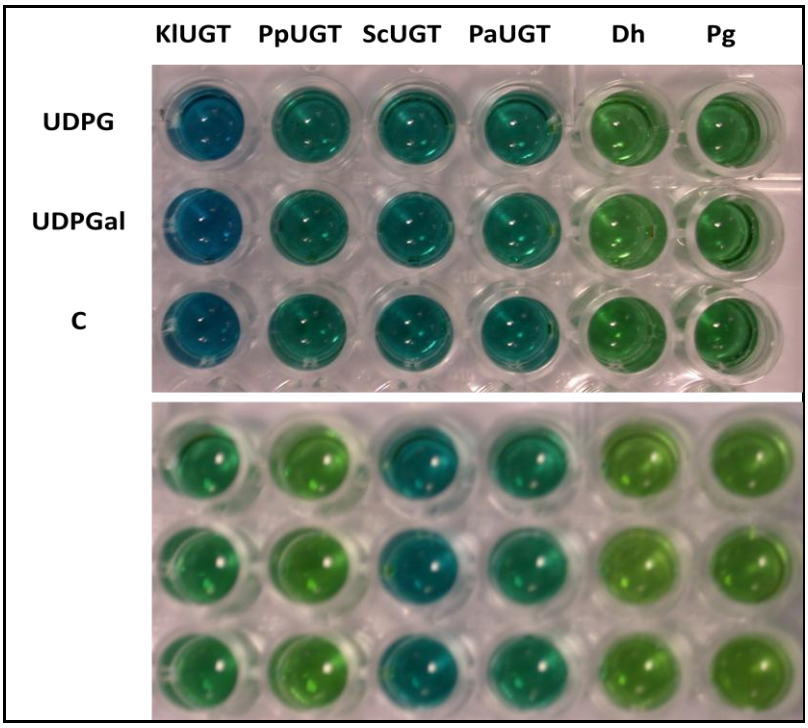


Fig. 4.3 Colorimetric assay screening using BB. Reactions were performed against cholesterol as acceptor and maintained at pH 7.3. A_{615nm} was measured at t=0 (above) and t=3 h (below) at 30^{o}C. C= control reactions where UDP donor was absent.

4.3.1.2 Screening ep-PCR library for enhanced UGT activity using UDP*G*

The ep-PCR library generated (**section 4.3.2**) for each clone was screened for significantly improved UGT activity compared to wt UGTs against cholesterol as the acceptor substrate using UDP*G* as donor. Cholesterol was chosen as the acceptor candidate for screening as all WT UGTs were active against this sterol which was not the case for testosterone.

Since transformation rate for HTS is critical, RM reactions were transformed in electrocompetent BL21 (DE3) (section **2.2.1.12**). This resulted in at least tenfold higher transformation efficiency (4000-5000 CFUs) of EZC reactions as compared to EZK reactions (100-350 CFUs) for each clone.

For each variant library, 94 individual colonies following transformation were picked, cultured and CFE extracts prepared using sterile 96-well deep well plates as described in section **4.2.4.2.** A similar screening approach has been recently undertaken to identify GT variants from an ep-PCR library with improved activity towards UDPG by using a 4-methylumbelliferone, a fluorescent acceptor (Williams and Thorson 2008). However, an advantage of the Williams and Thorson (2008) study was that the screening procedure could be entirely performed within a 96-well plate format since the quench in fluorescence following a GT reaction could be recorded using a plate reader. In my study however, the final stage of screening involved the extraction of radiolabelled products for LSC measurements which required assays using CFEs to be performed in tubes.

Tables 4.6 (A.-D.) depict the HTS screening data following LSC whereas **Table 4.7** summarises the HTS figures of cholesterol active mutants. As expected, due to the different mutagenic rates produced by the regimes followed, a maximum of only 10% of the hits screened from the Mutazyme generated variant library were active as compared to upto 45% active hits obtained from the Taq pol generated library of mutants. These results suggest that the high rate of mutation employed in the former case resulted in a majority of inactive enzymes. On the other hand, a mega-library was obtained with different mutagenic rates. CFE from pET-28a vector transformed into *E. coli* (BL21) served as a negative control and provided negligible background of less than 100 dpm in all cases which was in line with past studies where *E. coli* cell free extracts have shown no background activity in a radio-labelled assay format (Warnecke *et al.* 1999).

Sc mutant library	Dpm	fold increase
Mutazyme		
WtScUGT51'	505.86	
ScEZC24	**1177.9**	**2.3**
ScEZC35	465.19	
ScEZC46	**2197.1**	**4.3**
ScEZC82	**1789.21**	**3.5**
Taq pol		
ScEZK1	675.4	
ScEZK6	1136.77	
ScEZK7	869.15	
ScEZK11	877.25	
ScEZK12	1065.45	
ScEZK15	1145.4	
ScEZK16	775.28	
ScEZK17	599.54	
ScEZK19	500.26	
ScEZK20	**1166.14**	**2.3**
ScEZK21	501.93	
ScEZK23	971.13	
ScEZK24	731.23	
ScEZK27	627.48	
ScEZK28	983.31	
ScEZK31	1009.72	
ScEZK33	689.07	
ScEZK37	564.13	
ScEZK38	606.58	
ScEZK39	845.88	
ScEZK42	1103.5	
ScEZK44	952.95	
ScEZK45	762.41	
ScEZK47	**1686.54**	**3.3**
ScEZK48	926.19	
ScEZK50	553.83	
ScEZK53	932.59	
ScEZK56	426.75	
ScEZK57	784.61	

Table 4.6 A. HTS screening data for Sc mutant libraries. Hits showing at least two-fold increase in activity following LSC are indicated in bold and discussed in text.

Pp mutant library	Dpm	fold increase
Mutazyme		
WtPpUGT51'	797.28	
PpEZC8	336.36	
PpEZC12	339	
Taq pol		
PpEZK1	1304.96	
PpEZK6	1409.51	
PpEZK7	1492.46	
PpEZK9	1846.16	
PpEZK12	945.11	
PpEZK15	**4785.57**	**6**
PpEZK20	1139.43	
PpEZK22	1050.33	
PpEZK23	1475.67	
PpEZK30	**2380.25**	**3**
PpEZK36	820.9	
PpEZK46	**2477.69**	**3**
PpEZK50	847.24	
PpEZK58	950.14	
PpEZK59	1143.27	
PpEZK61	1625.13	
PpEZK66	1353.77	
PpEZK70	2129.7	
PpEZK75	1063.12	
PpEZK85	959.82	
PpEZK87	1196.96	
PpEZK89	748.6	
PpEZK93	573.6	

Table 4.6 B. HTS screening data for Pp mutant libraries. Hits showing at least two-fold increase in activity following LSC are indicated in bold and discussed in text.

Pa mutant library	Dpm	fold increase
Mutazyme		
WtPaUGT51'	2395.17	
PaEZC6	832.88	
PaEZC11	685.68	
PaEZC17	458.19	
PaEZC20	768.83	
PaEZC22	538.61	
PaEZC36	429.99	
PaEZC47	**6193.47**	**2.6**
PaEZC57	505.79	
Taq pol		
PaEZK2	1135.56	
PaEZK3	1032.59	
PaEZK4	875.14	
PaEZK5	721.73	
PaEZK6	529.5	
PaEZK8	1579.56	
PaEZK10	1774.08	
PaEZK13	2941.86	
PaEZK20	634.28	
PaEZK26	433.74	
PaEZK35	509.92	
PaEZK36	547.43	
PaEZK38	943.4	
PaEZK40	613.93	
PaEZK43	617.95	
PaEZK48	2158.22	
PaEZK49	570.28	
PaEZK50	1421.34	
PaEZK52	753.08	
PaEZK58	1798.91	
PaEZK61	1697.07	
PaEZK62	751.07	
PaEZK64	538.42	
PaEZK67	**4177.28**	**1.7**
PaEZK72	1266.27	
PaEZK73	2780.18	
PaEZK74	490.04	
PaEZK76	905.62	
PaEZK77	598.43	
PaEZK86	896.17	

PaEZK89	1042.82
PaEZK91	1205.56
PaEZK92	967.25
PaEZK94	1956.55
PaEZK95	1105.52
PaEZK96	3341.57

Table 4.6 C. HTS screening data for Pa mutant libraries. Hits showing at least two-fold increase in activity following LSC are indicated in bold and discussed in text.

Kl mutant library	Dpm
Taq pol	
WtKlUGT51'	782.83
KlEZC4	1037.2
KlEZC8	1180.44
KlEZC10	1106.75
KlEZC11	1216.48
KlEZC13	894.07
KlEZC14	1034.98
KlEZC15	1144.97
KlEZC17	1050.01
KlEZC18	625.34
KlEZC24	1203.44
KlEZC26	1755.46
KlEZC28	1207.04
KlEZC31	935.88
KlEZC33	1432.76
KlEZC34	1256.03
KlEZC35	1318.15
KlEZC39	1113.39
KlEZC40	734.43
KlEZC41	1336.31
KlEZC43	706.82
KlEZC47	941.2
KlEZC48	1223.83
KlEZC50	1415.12
KlEZC51	1628.34
KlEZC52	917.26
KlEZC56	899.4
KlEZC58	1412.81
KlEZC60	1430.08
KlEZC62	1083.78
KlEZC66	761.17
KlEZC68	842.03
KlEZC71	1345.09
KlEZC73	931.49
KlEZC75	1400.45
KlEZC76	1229.31
KlEZC77	970.11
KlEZC79	1176.32
KlEZC81	965.16
KlEZC85	1267.98

KlEZC89	1199.36
KlEZC90	1369.47
KlEZC91	755.1

Table 4.6 D. HTS screening data for Kl mutant libraries. Hits showing at least two-fold increase in activity following LSC are indicated in bold and discussed in text.

Mutant library	Mutazyme		Taq pol	
	Total screened	Hits	Total screened	Hits
ScUGT51'	92	4 (3)	59	28 (2)
PpUGT51'	56	2 (-)	94	23 (3)
PaUGT51'	77	8 (1)	94	36 (1)
KlUGT51'	63	-	94	42 (0)

Table 4.7 Comparison of active variants (hits) following ep-PCR with Mutazyme and Taq. Numbers within brackets represent the active mutants exhibiting at least double the amount of incorporation of UDP*G* as compared with wt UGTs as noted from the dpm values. Dpm values for the negative control were typically less than 100 dpm whereas wt clones transformed in BL21 demonstrated activity in all cases.

4.3.2 RM, transformation and generation of UGT variants

ep-PCR with Taq pol or Mutazyme was carried out with the aim of achieving different mutagenic rates. It has been previously shown that additional rounds of ep-PCR using Taq led to an increase in mutation frequencies in each successive round which was confirmed by the respective increase in thermostability of xylanases and loss of catalytic activity at a permissive temperature of 37^{o}C (Andrews *et al.* 2004). However, only one round of ep-PCR was performed when using Taq pol in my studies. However, when using Mutazyme, conditions for high mutagenic frequency (9-16 mutations/Kb) were employed.

Another crucial step when performing ep-PCR using Taq is the ligation of mutated gene fragments to the recipient vector that requires optimisation of the

ratio and concentrations of insert and vector (Hibbert and Dalby 2005; Miyazaki 2002). However, in my study, I developed a PCR cloning stage using a high fidelity KOD DNA polymerase (EZK reaction), analogous to the PCR cloning stage of the Genemorph II EZClone kit. Thus, MMs generated for WT clones (pScUGT51', pPpUGT51', pPaUGT51' and pKlUGT51') were purified and these served as primers to replace the homologous region in the template plasmid via the EZK and EZC reactions (section 4.2.3). The high fidelity DNA polymerases were used to prevent incorporation of new mutations during amplification of the entire r-plasmids. The resulting mixture was then treated with *Dpn*I in order to eliminate the original WT r-plasmids used as the donor template for PCR cloning reactions (Stage 2, **Fig 4.1**).

For preliminary analysis of mutation frequency, *Dpn*I-treated EZC and EZK reaction mixture for each library (2 sets of mutant libraries were generated for each wt clone) was transformed into chemically competent *E. coli* TOP10. To check the effectiveness of the EZC method, the supplied positive control plasmid containing the *lacZ* gene was used that resulted in a total of 535 CFUs of which 435 were white and the remainder blue when plated on agar media substituted with 80 µg/ml Amp, 240 µg/ml IPTG and 100 µg/ml X-gal. This suggested that approx. 81% of the total clones were phenotypic mutants even though the actual mutation rate may be much higher as not all mutations may lead to a change in phenotype. Two randomly picked clones from each variant library (i.e. 4 clones for each target) were verified to contain target inserts by restriction endonuclease digestion. DNA sequencing data from these revealed an average rate of 2-5 mutations/Kb for all target fragments generated by Taq and a relatively higher rate

of 3-9 mutations/Kb target fragment generated by Mutazyme. A low mutation frequency may not necessarily lead to an amino acid whereas several mutations could lead to a majority of inactive enzymes. Mutation rate (transitions or transversions) can be further increased by lowering the target DNA for ep-PCR to as low as 0.1 ng which is proposed to yield the highest mutation frequency. It has been previously shown that a mutagenesis frequency corresponding to 1-2 amino acid changes per target per round of ep-PCR using either Taq or Mutazyme was preferred in order to prevent accumulation of neutral mutations (Bulter *et al.* 2003).

Variants displaying the highest level of UGT activity as well as a few random variants displaying at least a two-fold increase in rate of incorporation with UDP*G* as compared with the corresponding parental enzyme were selected for mutational analysis. For this purpose, the extracted pDNA was transformed into chemically competent *E. coli* TOP10 cells prior to sequencing reactions. Variants with enhanced levels of UGT activity compared to native enzymes, listed in **Table 4.8** were found to contain substitutions in both the *N*- and *C*-terminal regions. Each variant is referred to by the respective organism abbreviation as the prefix followed by the method used to generate the mutant library followed by the unique number for the clone, for example ScEZC46.

Variant	Mutation	Activity (relative to native UGTs)
ScEZC46	E61V	two-fold
ScEZC24	G189E	Similar
PaEZC62	V316A E433G	six-fold
PpEZK15	K207R N400S	six-fold

Table 4.8 UGT variants with improved activity compared to respective native enzymes following RM.

4.3.3 Structural homology and analysis of UGT variants

4.3.3.1 Homology modeling and overall structure of ScUGT51'

When used as a query sequence to match 3-D structures within the RCSB Protein Data Bank (http://www.pdb.org/pdb), ScUGT51' showed closest structural homology with family 1 glycopeptide antibiotic GTs (Gtfs) from *A. orientalis* giving *E*-values of $1.49x10^{-22}$, $8.37x10^{-20}$, $2.69\ x10^{-18}$, and $3.0x10^{-14}$ against a chimeric GtfA (GtfA'; PDB id: 3H4I) (Truman *et al*., 2009), GtfB (PDB id: 1IIR) (Mulichak *et al*., 2001), GtfA (PDB id: 1PN3) (Mulichak *et al*., 2003) and GtfD (PDB id: 1RRV) (Mulichak *et al*., 2004), respectively. Primary sequences of all the Gtfs derived from the PDB database exhibited an overall 62% identity among them and ScUGT51' displayed a 25-28% overall sequence identity to GtfA', GtfA, GtfB and GtfD. GtfD and GtfA utilise vancosamine and *epi*-vancosamine sugars for vancomycin and choloroeremomycin biogenesis respectively whereas GtfB is specific for UDPG in choloroeremomycin biogenesis and utilise different

NDP-sugars and acceptor aglycones. GtfA and GtfD are however, structural homologs and decorate the same monoglycosylated vancomycin intermediate-desvancosaminyl vancomycin (DVV) (Walsh *et al.* 2003).

ScUGT51' was modelled based on the crystal structure of GtfA, GtfA', GtfB and GtfD as templates to gain further insight into substrate binding and catalytic mechanism. The secondary structure of ScUGT51' was predicted by PSIPRED protein structure prediction program (UCL, http://bioinf.cs.ucl.ac.uk/psipred/) and was used as a guide to produce an optimal structural alignment with Gtfs (**Fig. 4.4.**). For this purpose, the ScUGT51' sequence was truncated at the *N*- and *C*-terminus (final sequence: ^{18}Y-Y^{439}) to fit the template sequences. Molecular modeling was performed by Dr. Marcus Durrant, Northumbria University by using DeepView/Swiss-PDB Viewer (version 4.0.1; available at http://www.expasy.org/spdbv) to construct a valid model of ScUGT51' based on its optimally produced alignment. Loops were built using the DeepView loop database. Energy was minimised to 1 Kcal/Å in Hyperchem using AMBER force field. The ligands thymidine diphosphate (TYD) and DVV of the GtfD-complex (PDB id: 1RRV) were used for ligand docking in the ScUGT51' model.

GtfB, GtfA and GtfD are all GT-B fold members with a bi-lobed 3-D structure adopting a GT-B fold (Mulichak *et al*. 2003; Mulichak *et al.* 2001; Mulichak *et al.* 2004). The ScUGT51' model based on the structural alignment is shown in **Fig. 4.5.** Consistent with the structures of the Gtfs (Mulichak *et al.* 2001)., the structure of ScUGT51' also belonged to the typical GT-B fold where both the *N* and *C*- terminal domains consist of a core structure of parallel β-sheets numbered as Nβ1-7, Cβ1-5. In addition, as in the case of other GT-B fold GTs, a long helical

tail of three consecutive helices is present at the end of the *C*-terminus and is shown to be involved in hydrophobic interactions linking the two domains (Wang 2009). GT-B fold-characteristic glycine-rich motifs such as those harboured within the USS, are associated with donor sugar binding and are present in all sequences as shown. Based upon these observations, all UGTs in this study may therefore adopt a similar core structure.

Flexible loop regions in the 3-D structure of all Gtfs correspond with hypervariable sequence regions (**Fig. 4.4.**) and are responsible for the optimal binding positioning of the vancomycin aglycones (Mulichak *et al.* 2001). Only the GtfA, GtfA' and GtfD structures were available in NDP and/or aglycone bound forms where the conformation as expected was cleft-closed which allows the *N*-terminus N3 and N5 loops (**Fig.4.5**) following the Nβ4 and Nβ5 respectively to form the acceptor binding crevice as well and promotes donor sugar interactions with both domains. On the other hand, both the N3 and N5 loops were disordered in the ScUGT51' model similar to the non ligand-bound GtfB structure further confirming that these are involved in interactions with the acceptor.

GtfD has more elaborate interdomain contacts where specific residues within the flexible loops have been shown to interact with the *C*-terminal domain (Mulichak *et al.* 2004).

As the interdomain linker is immediately followed by a helix in Gtfs that also corresponds to a helix in the predicted secondary structure of ScUGT51', this suggests that the position of the linker may be similar in ScUGT51'. However, the length of the interdomain linker is variable in plant UGTs and Gtfs, where it is much shorter in the latter. Differences in length and composition of linker region

serve to define domain positioning and substrate accommodation (Osmani *et al.* 2009).

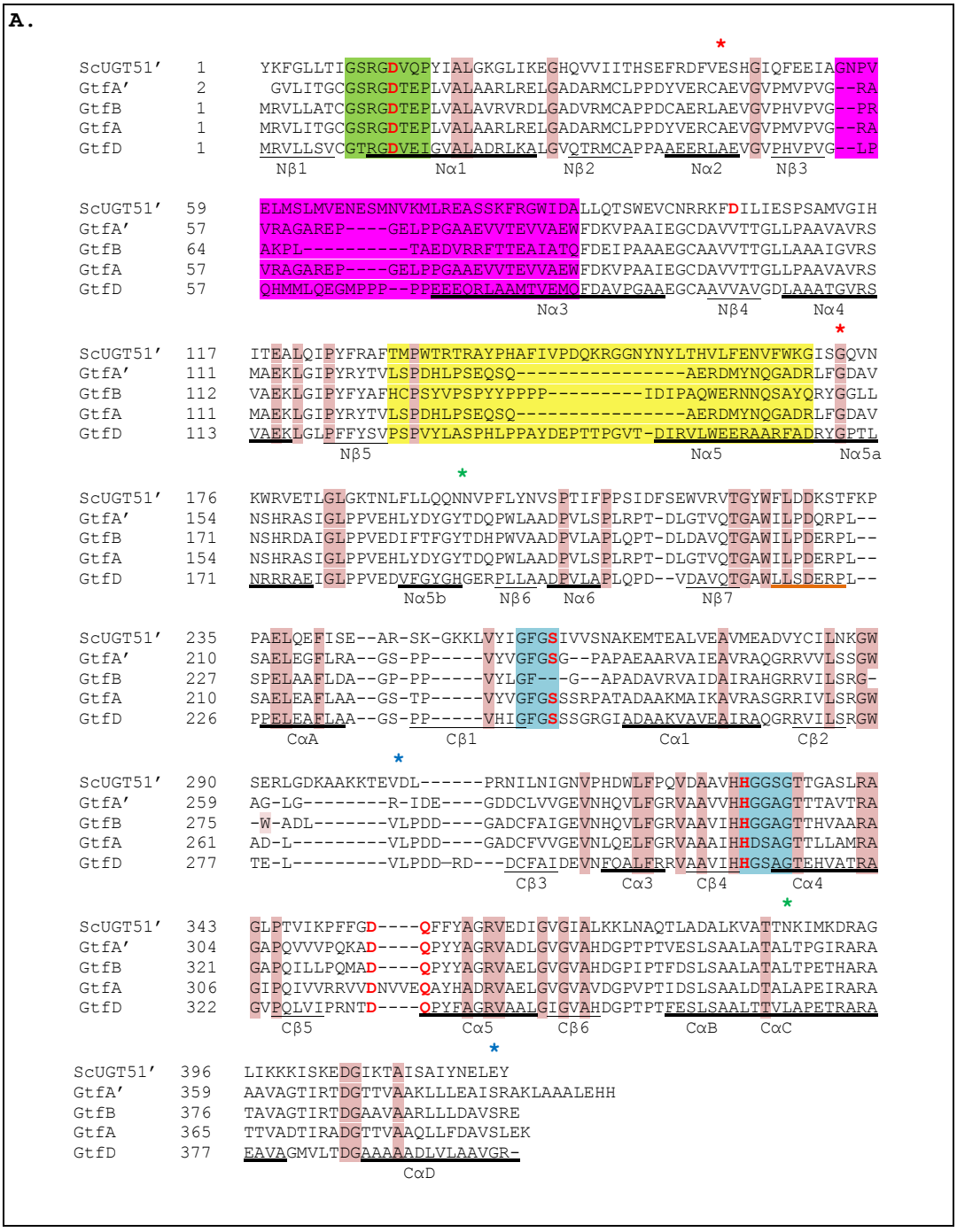


Fig. 4.4 Structure-based sequence alignment of ScUGT51' and Gtfs. The secondary structure elements of GtfA (PDB id: 3H4I); GtfB (PDB id: 1IIR); GtfA' (PDB id: 1PN3) and GtfD (PDB id: 1RRV) are indicated below the alignment and underlined. Conserved residues (pink), glycine-rich regions in the *C*-terminal (blue) and *N*-terminal (green) are highlighted. Catalytic residues (discussed in text) are shown in red. Interdomain linker peptide is underlined (orange). Disordered N3 (magenta) and N5 (yellow) loop regions in Gtfs and corresponding sequence elements in ScUGT51' are highlighted. Stars above the

sequence represent locations of substitutions in ScUGT51' (red), PaUGT51' (blue) and PpUGT51' (green) mutants.

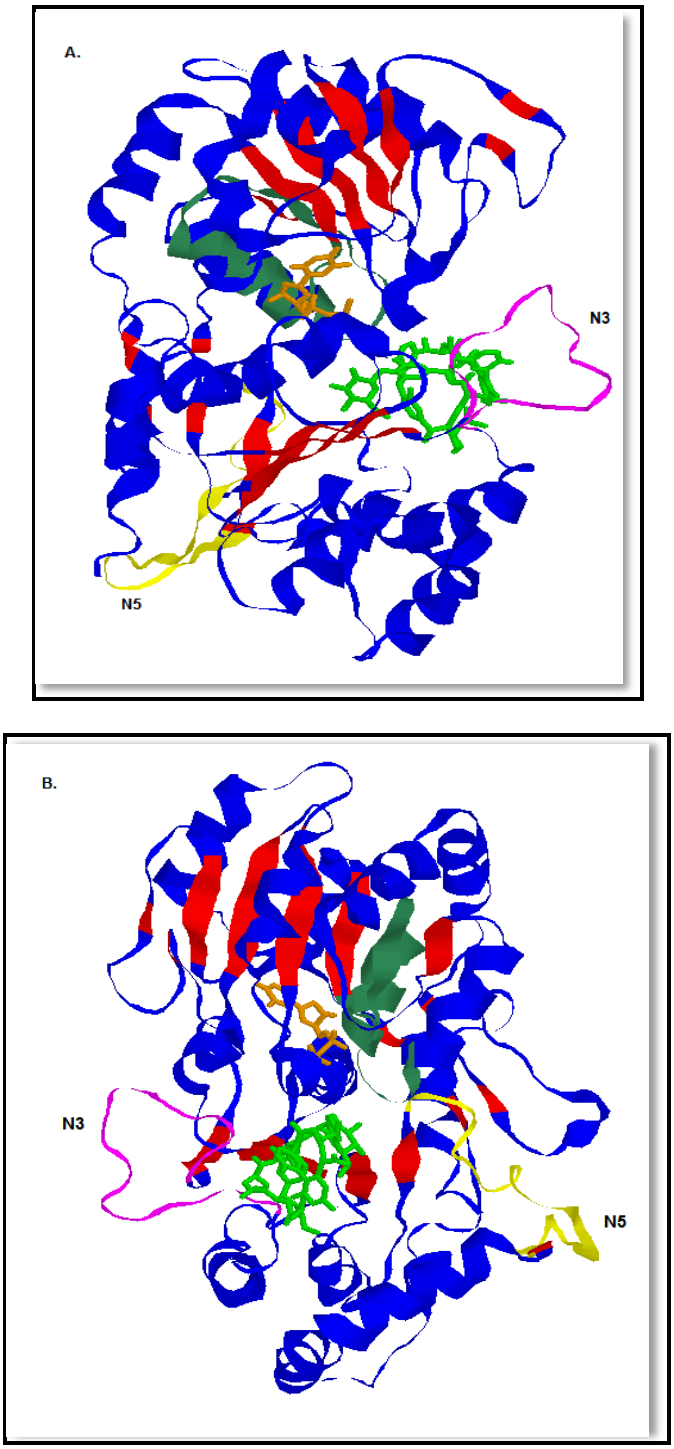


Fig. 4.5 ScUGT51' model and fitted DVV and TDP. A., B.: Different views of the structure showing the long *C*-terminal helical tail joining the *N*-terminus in **A**.DVV (green) and TDP (orange) ligands fitted from the superposition with GtfD:DVV:TDP complex (PDB id: 1RRV) are displayed as stick models. Hypervariable loops N3 (pink), N5 (yellow) are shown. The USS is coloured in grey. The figure was prepared using RasMol.

4.3.3.2 Acceptor and NDP-sugar interactions

The ligand docking study and comparison with previously reported GT-B structures indicated that D13, S10, D103, S259, H330, D354 and Q355 were the key catalytic residues of ScUGT51'. The catalytic site harbouring these residues is shown in **Fig. 4.6.**

The general catalytic base

D13, corresponding to D13 in ScUGT51' has been proposed as the catalytic base of glycosyl transfer in the case of GtfA, GtfA' and GtfD (Mulichak *et al*. 2003; Truman *et al.* 2009; Mulichak *et al.* 2004). The D35 of HMT31, another chimeric antibiotic GT (Park *et al*. 2009b) was matched with the D13 as a catalytic general base and in fact the eight residue glycine-rich motif in the *N*-terminus (GS(T)RGDVQP) containing the catalytic base is common to many other putative and characterised sequences of GT-B members from plants, other fungi and viruses (Hacker et al. 2005) suggesting that it consists of residues important for glycosyl transfer. This also includes the highly conserved S/T10 residue that is involved in hydrogen bonding with the reactive –OH of GtfA and GtfD (Mulichak *et al.* 2003; Mulichak *et al.* 2004). However mutational analysis of this D13 in GtfB revealed that although the k_{cat} displayed by the mutant reduced ten-fold, the K_m value for UDPG remained unaffected. D332 (corresponding to D354 in ScUGT51') was proposed as the catalytic base instead. On the other hand, the D13A GtfD mutant showed a 5500-fold drop in k_{cat} confirming its role as the catalytic base. Moreover, its proximity to DVV as well as TDP strongly suggests this as the base. In plant UGTs however, for example UGT71G1 and UGT73A5, the corresponding Asp position is occupied by H22 which is close to the C1' of

the bound UDPG (Hans *et al.* 2004). Also in this case, an acidic residue D121 (corresponding to D103 in ScUGT51') was proposed to be involved in forming an electron transfer chain to H22/D13 to assist deprotonation of acceptor (Shao *et al.* 2005). It has been shown that the interaction of H22 and D121 is essential for enzyme activity (Brazier-Hicks *et al.* 2007). Interestingly, even though this Asp residue is absent in all Gtfs, the D103 residue in GtfD (corresponding to Leu in other Gtfs) has been suggested to create a hydrophilic catalytic pocket.

Acceptor binding site

Acceptor interactions are largely as a result of the distinct patches of hydrophobic and aromatic residues in the N3 and N5 loop regions that are involved in acceptor interactions (Mulichak *et al.* 2001). Interestingly however, the sequence and length of these loops greatly varies among other Gtfs, such as these loops in GtfD are much longer compared to GtfA and GtfB. ScUGT51' although lacks the presence of hydrophobic residues involved in such interactions, the loop regions are similar in length to GtfD. On the other hand, certain interdomain interactions are conserved across Gtfs such as that of R11 and S230 in GtfA (corresponding to R11 and S259 in ScUGT51'). Also, in addition to positive helix dipole interactions suggested for GT-B fold GTs, other residues in Gtfs are shown to be involved in ionic interactions to stabilise the thymine base of TDP such as that between R11 and E277 of GtfA and are not likely to be present in ScUGT51' due to the presence of a conserved Asn residue in all yeast UGTs corresponding to E277. In this case positive helix dipole interactions of the Nα1 and Cα4 could contribute to hydrogen bonding interactions promoted by the highly conserved glycine residues across distantly related GT-B members.

It has been suggested earlier that acceptor selectivity could be altered in GT-B members by targeting the loop between Nβ5-Nα5 in the acceptor binding domain. This loop was conserved among MurG homologs that utilise the same acceptor and corresponded to the variant loop in the case of Gtfs that utilise subtly different acceptors despite high sequence homology (Hu and Walker 2002). The sequential, ordered reaction mechanism could also influence substrate recognition and binding. For example in the case of MurG, the NDP-sugar binds first whereas in the case of the related macrolide OleD, the acceptor binds prior to the NDP-sugar (Quiros *et al.* 2000). In Gtfs however, the order of substrate recognition is still unclear as for GtfB it was suggested that the binding of the UDPG ligand required prior aglycone binding whereas an independent substrate binding mechanism has been suggested for GtfA.

NDP-sugar binding site

In the ScUGT51' model, the donor sugar moiety resides within the inter-domain cleft as has been earlier shown for GT-B structures. The closed confirmation of GT-B GTs allows for the pyrophosphate moiety of the NDP-sugar to interact simultaneously with the *N*- and *C*-terminal domains and therefore requires the prior binding of the NDP moiety (Mulichak *et al.* 2003).

H330 and G334 of ScUGT51' are part of the ^{329}HHXXAGT^{335} loop which is a highly conserved glycine-rich motif among distantly related GT-B fold GTs (**Fig. 4.7)** constituting a large part of the donor binding site ((Hu *et al*. 2003; Hu and Walker 2002; Rix *et al.* 2002). For example it is a binding site of α-phosphate of the NDP-sugar donor in GtfA (Mulichak *et al.* 2003). UDP-glucuronosyltransferases that are involved in the glucuronidation of various

natural as well as foreign compounds including carcinogens and drugs also contain the consensus USS ((Mackenzie *et al.* 1997; Meech and Mackenzie 2010).

The Q289 of MurG (corresponding to Q355 in ScUGT51') located within this loop of the β/α/β motif was proposed to hydrogen bond with the -OH at the C-4 of the sugar which suggests that the isomer galactosamine may not be accepted. (Hoffmeister *et al.* 2001; Hu and Walker 2002) This is because the axial O-4 of UDPGal has been proposed to create steric hindrance with the acceptor oxygen even though it may fit into the active site (Thorsoe *et al.* 2005). Interestingly, as shown in the alignment in **Fig. 4.7**, GTs utilising UDPG or UDP-GlcNAc as the sugar donor have Gln as the conserved sugar recognition determinant whereas in the ceramide GalT (Kapitonov and Yu 1999), His occupies this highly conserved position. In one study, H374 of a plant GalT (corresponding to the highly conserved Q382) upon being replaced by Gln switched galactose preference to glucose (Shao *et al.* 2005). However, the glucose specificity for a flavonoid GT (UGT71G1) could not be altered when Q382 in this case was mutated to His, suggesting that other residues may be additionally play a role in dictating donor specificity (Shao *et al.* 2005). This study also showed that docking of UDPGal in the structure of UGT71G1 would prevent formation of hydrogen bonds with the adjacent E381, another key conserved residue, and therefore suggestive of a multi-residue involvement in sugar recognition and specificity. In addition, this residue is conservatively replaced by Asp in several eukaryotic GT-B GTs including the UGTs in the present study, further highlighting its role in NDP-sugar recognition.

A few conserved residues within the USS across GT-B members are directly involved in forming hydrogen bonds with the NDP-sugar moiety. Thorsoe *et al.*

(2005) established by mutational analysis of UGT85B1 that certain residues within the USS motif were critical for catalysis. For example the C-6 of UDPG hydrogen bonds with E410, Q411 (corresponding to D354 and Q355 in ScUGT51'). Interestingly, E410 in this case has also been shown to hydrogen bond with both the acceptor and donor which has not been reported in the case of Gtfs. Another critical hydrogen bonding interaction in this case involves the S391 of UGT85B1 that interacts with pyrophosphate of donor. This residue corresponds to a conserved Thr residue among Gtfs (for example T296 of GtfA), as well as T335 of ScUGT51'. The conserved role of the Glu and Gln residues has also been confirmed in the case of other plant UGTs such as the arbutin synthase (AS) (Hefner *et al.* 2002).

In the case of UGT71G1 (Shao *et al.* 2005), another plant UGT, it was shown that the phosphate of the sugar forms hydrogen bonds with H357 (corresponding to H330 in ScUGT51) and S285 (corresponding to S259), the only residue outside the USS which is directly shown to interact with the phosphate group as shown in **Fig. 4.6**.

The nucleotide pocket environment however, differs between the plant UGTs and the Gtfs (Offen *et al.* 2006). Apart from the residues within the USS involved directly in NDP-sugar binding, other residues within the PSPG motif of plant sequences are shown to have been involved in creating a hydrophobic environment around the donor molecule. For example. the indole ring of the conserved tryptophan (W), the first residue of the PSPG motif is involved in donor recognition and hydrophobic interactions in all plant UGT structures solved so far (Osmani *et al.* 2009).

In UGT71G1 binding UDP-glucose, MurG binding UDP-GlcNAc, GtfD binding TDP, sugar donor molecules are positioned similarly despite low sequence identities and detailed interactions with sugar donors also vary.

Interestingly, some residues present in vancosamine recognising Gtfs differ from those recognising UDPG as the donor. For example, hydrophobic residues (P126 in GtfD and L124 in GtfA) favourably interact with vancosamine whereas a substituted H125 in GtfB was proposed to form H-bonds with UDPG. A T130 residue in the ScUGT51' may also offer potential H-bonding interactions with the 6'–OH of UDPG. Also, S311 in GtfD involved in sugar binding is conserved in GtfA and is substituted with glycine in the case of GtfB and ScUGT51' strongly suggesting its role in the exclusive recognition of UDPG.

Residues involved in dictating catalytic mechanism

Differences in specific residues among GT members may lead to difference in catalytic mechanism. For example, as described earlier, a His residue is located in the first position of the USS of GT-B fold GTs in the case of family 1 and family 28 members. However, this position is occupied by Asp or Glu residues for GT family 4 and 20 members that catalyse sugar transfer by the retaining mechanism. Consequently, charge differences in these usually well conserved motifs may dictate catalytic mechanism (Rosen *et al.* 2004).

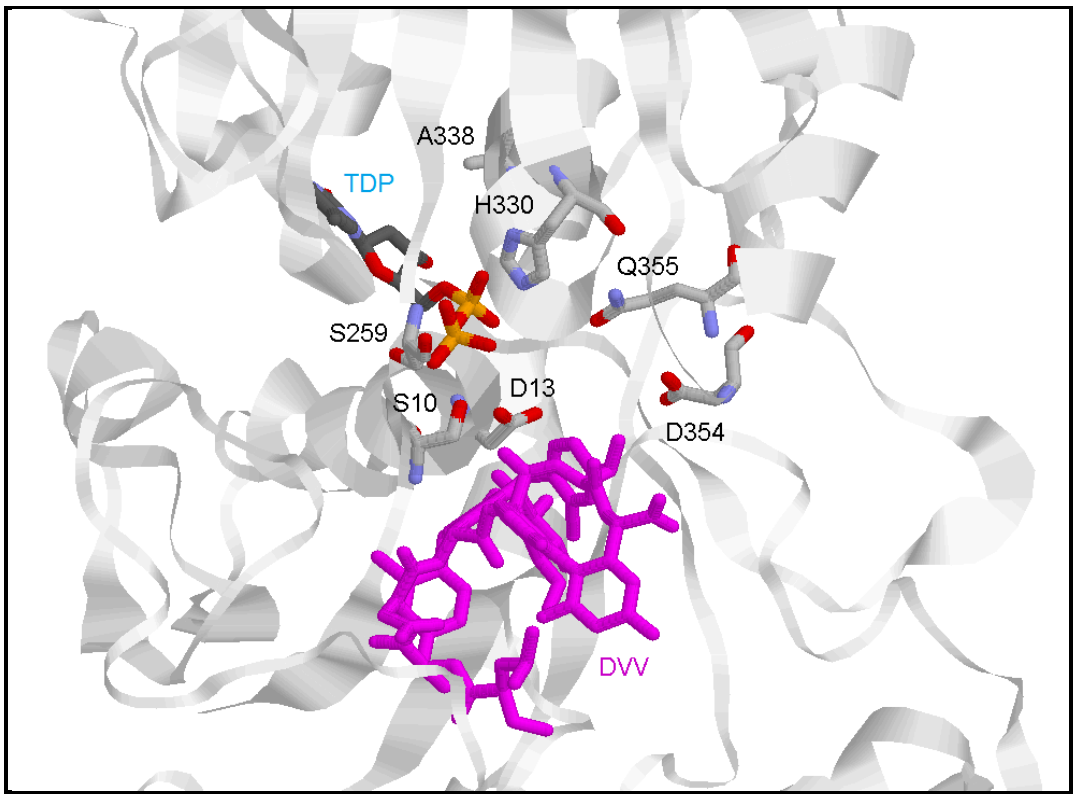


Fig. 4.6 Schematic showing significant active site residues and fitted DVV and TDP ligands in the ScUGT51' model. DVV and TDP (fitted from the superposition with GtfD: DVV: TDP complex; PDB id: 1RRV) are indicated and proposed residues involved in glycosyl transfer reactions are shown as CPK models. The figure was prepared using RasMol.

```
UGT3A              366 HGGQNSIMEAIQHGVPMVGIPLEG---DQPENMVRVEAKKFGVS
AS                 360 HCGWNSILESVVNGVPLIAWPLYA---EQKMNAVMLTEGLKVAL
OleD               306 HAGAGGSQEGLATATPMIAVPQAV---DQFGNADMLQGLGVARK
GtfB               308 HGGAGTTHVAARAGAPQILLPQMA---DQPYYAGRVAELGVGVA
UGT51'             347 HGGSGTTGASLRAGLPTVIKPFFG---DQFFYAGRVEDIGVGIA
cGalT              358 HGGLNSIFETMYHGVPVVGIPVFG---DHYDTMTRVQAKGMGIL
MurG               261 RSGALTVSEIAAAGLPALFVPFQHKDRQQYWNALPLEKAGAAKI
                       :.*  .       . * :  *      ::      :   :.
Consensus              H_G__T__E____G_P____P_____DQ___A_____G_G__
```

Fig. 4.7 Sequence alignments of the USS of GT-B fold family members. Dots represent similar residues and asterisks represent identical residues in all sequences. The consensus sequence is shown below. The sequences referred to are: UGT3A (human UDP-GlcNAc GT), AS (arbutin synthase, *R. serpentina*), OleD (oleandomycin GT, *S. antibioticus*), GtfB (chloroeremomycin GT, *A. orientalis*), UGT51' (yeast UGTs in this study), cGalT (mammalian ceramide GalT) and MurG (UDP-*N*-acetylglucosaminyltransferase, *E. coli*).

4.3.3.3 Analysis of UGT variants

As mentioned earlier, point mutations within the *N*- and/or *C*-terminal domains led to increase in GT activity. In the case of the ScEZC24 mutant (G189E), a highly conserved glycine in an *N*-terminal helix when substituted with a negatively charged glutamic acid residue resulted in no significant increase in GT activity. However, in the ScEZC46 mutant (E61V), replacement of glutamic acid with a hydrophobic valine residue resulted in a two fold increase in GT activity as compared with the native enzyme. Interestingly, this position is occupied by hydrophobic alanine residues in Gtfs suggesting that hydrophobicity in this region may play a role in acceptor binding. The location of this residue in the 3-D model suggests that it is not involved directly in interaction with the acceptor and as such its role remains unclear. However as shown in the past, beneficial mutations remote from the catalytic site may influence substrate specificity, selectivity as well as catalytic activity of GTs (Williams and Thorson 2008; Williams *et al.* 2007).

In the case of the PaEZC62 double mutant (V316A; E433G) both mutations within the *C*-terminal domain resulted in a six fold increase in GT activity compared with the native enzyme. A hydrophobic Val was replaced by another hydrophobic residue, Ala. A Val residue at position 316 in a probable loop region is unique to PaUGT51' and corresponds with an Asp or Glu residue in all other UGTs; for example, it corresponds to D321 of ScUGT51' (**Fig. 3.2 C**). Interestingly, this Asp residue is also common to all Gtfs in a similar position in the *C*-terminal loop and there is no evidence of its interaction with the NDP-sugar. The second of these mutations, E433G (corresponding to E439 of ScUGT51')

may although play a synergistic role with the V316A, KlUGT51' also consists of glycine at the corresponding position. Moreover, this residue is present in the long helical tail that is not involved in donor recognition or binding.

Another double mutant, PpEZK15 displaying six-fold GT activity compared to the wt PpUGT51' also consists of a replacement (N400S) in the *C*-terminal long helical region that corresponds with N407 of ScUGT51'. The other substitution in this mutant, K207R in a short (3 residue long) *N*–terminal loop corresponding to N214 of ScUGT51' may play a role in substrate positioning within the catalytic cleft formed between the two domains. This is because interestingly, PpUGT51' consists of a Lys residue with a positively charged side chain at this position which was replaced by a similar residue, Arg in PpEZK15. Since PpUGT51' displays the highest affinity towards cholesterol compared to all other UGTs and this position is occupied by a polar, uncharged Asn residue in other UGTs, this suggests that this residue may be involved in acceptor binding. Moreover, the location of this residue within the acceptor binding crevice between the N3 and N5 loops further suggests that it may have a possible catalytic role.

4.4 Discussion

RM and functional analysis of UGTs

Mutagenesis approaches have served as a powerful tool for studying GTs with respect to substrate specificity and providing critical structural information. Evolving GTs with the desired substrate specificity for the synthesis of several oligosaccharides and complex carbohydrates for example has facilitated the development of potential therapies (Aharoni *et al.* 2006). On the other hand, enhancing activity towards natural, non-natural substrates serves as the most popular starting point for mutagenesis studies (Dalby 2003). Family 1 GTs are attractive subjects for several DE studies based on the already existing knowledge regarding substrate specificity and catalytic activity. The crystal structures of 58 GT-B fold members have been determined so far of which 19 have been assigned to GT family 1, the largest structural database for any GT family. Moreover, a majority of sequences within this family lack TM segments and therefore are amenable to cell free assays for screening large libraries of variants. Despite existing knowledge of this family of GTs, little is known in terms of the function of family 1 USGTs. The aim to evolve USGTs with higher catalytic efficiency apart from improving SG synthesis would allow a better understanding of USGT function.

However, the success of a directed evolution experiment depends primarily on the sensitivity and throughput of the screen (Zhao *et al.* 2002). Thus it is crucial to optimize the screening procedure for these studies. (Palcic and Sujino 2001) have extensively reviewed various assays amenable to GTs. Some examples include

immunological assays that require specific antibodies for product identification and chromogenic and fluorogenic assays that require substrate or donor fluorescence or labelling coupled with special detection requirements not suited for high-throughput screening purposes. Coupled spectrophotometric assays based on the detection of UDP by coupling its synthesis to NADH oxidation with pyruvate kinase and lactate dehydrogenase have also been previously reported (Gosselin *et al.* 1994; Jackson *et al.* 2001). However, these assays cannot be used for screening crude cell extracts due to interference by NADH oxidases (Persson and Palcic 2008). The FACS (fluorescence-activated cell sorter) technique is currently emerging as a rapid and sensitive screening system using whole cells (Aharoni *et al.* 2006). Although this approach requires the fluorescence labelling of either substrate, it alleviates the need for time consuming cell lysis steps and enables sorting of a large library (10^6) of variants. This technique also offers the advantage of assessing GT activity and substrate specificity simultaneously by using differentially labelled substrates within the system and sorting the resulting cell population consisting of fluorescently labelled products (Hibbert and Dalby 2005).

A simple colorimetric method based on monitoring the colour change which is directly proportional to proton release associated with a GT catalysed reaction was attempted in this study. Several such studies have been reported in the past for nitrilases, kinases and hydrolases which require an appropriate buffer/indicator combination of similar pK_a values (Janes *et al.* 1998). However in most cases, assay conditions such as buffer strength and substrate concentration require optimisation for accurate measurements that can overcome small pH changes

during the course of the assay. For example Banerjee *et al*. (2003) screened for nitrilase activity in crude cell suspensions using relatively high buffer (10 mM) and substrate amounts (12.5 mM). It was found that the relatively high buffer concentration reduced sensitivity but showed a more linear relationship between absorbance and proton concentration. Persson and Palcic (2008) also confirmed decrease in sensitivity of GT activity screening when acceptor amounts of less than 0.2 mM were used.

pH-sensitive assays have been only recently exploited for HTS and in the case of GTs a few examples of previous studies are summarised in **Table 4.9.** Deng and Chen (2004) first developed such an assay for 1, 4-GalT using PR as indicator; however this has not been adapted to microtitre plates for HTS. An agar plate based colorimetric assay has also been described to monitor GT activity in whole cells using cresol red (Park *et al.* 2009a). However due to longer incubation periods required to detect activity, diffusion of hydrogen ions to surrounding colonies may lead to false positives and therefore low sensitivity ((Ben-David A *et al.* 2008). HTS for saturation mutagenesis created blood GalT variants has been successfully described (Persson and Palcic 2008). BB was used as the indicator as its p*K*a value of 7.1 was very similar the optimum pH of 7.0 for GT activity in this case.

In order to realise the potential of colorimetric assays, further studies need to be undertaken with UGTs under optimised conditions to accurately translate increasing proton concentration to colour/absorbance change. Screening for improved activity leads to loss of substrate selectivity and vice versa (Ben-David A *et al*. 2008). Some reports have also argued that no obvious change in

absorbance associated with glycosidic bond formation and that screening still represents a major challenge (Aharoni *et al.* 2006). In this regard, a semi HTS system was developed in this study based on the sensitive and reliable assays using radio-labelled glucose (**Tables 4.6 A.-C**). This approach not only offered the advantage of isolating positive mutants, but also provided basis for screening UGTs displaying significant activity compared to the parental enzymes. Although the activities in the crude lysates were significantly lower than those determined for the purified enzymes, the sensitivity of the radiochemical assay was able to readily distinguish between active and inactive enzymes in a relatively large-scale screening system. However at present, due to the lack of availability of specific radio-labelled substrates and the huge costs associated with the synthesis and disposal of such compounds, this approach cannot not be fully exploited for the synthesis and screening of complex carbohydrates (Zea and Pohl 2004).

GT	cell based/CFE	Buffer (mM)	Acceptor (mM)	Reference
1,4-GalT	CFE	2	10	(Deng and Chen 2004)
antibiotic GT	cell based	NI	2	(Park et al. 2009a)
blood GalTs	CFE	2.5	0.5	(Persson and Palcic 2008)

Table 4.9 Examples of pH colour assays discussed in text. NI= not indicated.

Structural homology of GT-B fold GTs

HM is a highly valuable tool for retrieving information on topology, identifying important structural features such as specific residues involved in catalysis and for predicting the preference to alternative sugar donors (Osmani *et al.* 2009). It also represents a means to guide rational design and HM studies have been successfully verified in the past by biochemical analysis of wt GTs and mutants created based on information gathered by models (Thorsoe *et al.* 2005). ScUGT51' was chosen as the subject for HM as SG synthesis is a well documented phenomenon in *S. cerevisiae* as well as related yeasts, such as *P. pastoris* and therefore represents a platform for DE experiments in related GTs.

The ScUGT51' model was produced based on Gtfs that were the best available candidates in the PBD database. Gtfs belong to glycopeptide (group III) antibiotics, the others being polyketide (group I) that includes OleD and aminoglycoside (group II) (Park *et al.* 2009b). Glycopeptide antibiotics are used to treat infections predominantly caused by Gram-positive bacteria such as *Staphylococcus aureus.* Sugar moieties are added to different sites on similar aglycone substrates during the late stages of the assembly of the peptide antibiotics (Hu and Walker 2002). Despite high primary sequence homology (~60%), Gtfs utilise different donors as well are highly regiospecific. For example GtfD and GtfA utilise vancosamine and *epi*-vancosamine sugars for vancomycin and choloroeremomycin biogenesis respectively whereas GtfB is specific for UDPG in choloroeremomycin biogenesis and all Gtfs decorate the same monoglycosylated vancomycin intermediate- DVV (Walsh et al. 2003).

Comparison of the homology generated model of ScUGT51' and several other GT-B fold GTs revealed high structural similarities despite low overall sequence identity. The ScUGT51' model possess a similar NDP-sugar binding motif like other GT-B members such as Gtfs and MurG and in general, share a higher similarity in their *C*-terminal domains as compared to the *N*-terminal domains, suggesting a probable common evolutionary origin. The high conservation of the α/β/α motif across GT-B members suggests that donors are held in a similar manner, regardless of the nature of the acceptor. This was further confirmed by ligand docking studies, which placed TDP in a position to adequately interact with the conserved residues of the conserved *C*-terminus loop.

ScUGT51' showed the closest structural similarity to a chimeric GtfA' with a position specific scoring matrix *E*-value of $1.49x10^{-22}$.This GtfA' was created using the *N*-terminal domain of GtfA and the *C*-terminus of an undisclosed ORF and interestingly, it was found that the *C*-terminal domain was highly similar to that of the parent GtfA despite the promiscuity in the range of substrates displayed by the chimera (Truman *et al.* 2009). However, as opposed to the vancosamine sugars utilised by the parent GtfA, GtfA' showed specificity for UDPG and UDP-GlcNAc donors instead and that may explain its higher structural similarity to ScUGT51'. In fact the E-value for GtfA' was higher than that observed for GtfB, that has a strict UDPG preference (Mulichak *et al.* 2001).

Wang *et al*. (2008) have recently determined the solution structure of Alg13, the NDP-sugar binding domain of a family 1 UDP-GlcNAc GT from *S. cerevisiae*. It was found that this shared structural topology with the donor binding domain of family 28, GT-B UDP-*N*-acetylglucosaminyltransferase (MurG) from *E. coli*

which binds *N*-acetyl glucosamine in a *C*-terminal pocket for transfer to a lipid-linked acceptor for peptidoglycan synthesis (Hu and Walker 2002). Family 1 GTs are related to family 28 members in the overall CAZy system of classification, mainly due to the similarity in the *C*-terminal domains (Rosen *et al.* 2004). Interestingly however, even though GtfB shares a conserved topology with MurG (Mulichak *et al.* 2001), ScUGT51' did not show significant structural similarity with MurG. This may be explained by the fact that Gtfs possess different donor specificities by virtue of only a few primary sequence residues whereas MurG is highly specific for its donor sugars. For example GtfB utilises UDPG initiating the choloroeremomycin pathway, whereas GtfA and GtfD utilise vancosamine sugars for antibiotic assembly and therefore certain sequence elements of these GTs may overlap with those of ScUGT51' as well as MurG. Also, as shown in **Fig. 4.4,** instead of a conserved Asp residue in the case of other GT-B fold members, this position is occupied by Gln at position 288 in MurG which may further highlight its role in substrate specificity. Furthermore, in the case of MurG, the glycine-rich motif in the *N*-terminal domain does not make interactions with the NDP as opposed to Gtfs (Mulichak *et al.* 2004). Since the same residues are proposed to interact with NDP in the case of ScUGT51', this further suggests the similar catalytic mechanism between Gtfs and ScUGT51'.

Therefore, comparing enzymes with different selectivity but conserved primary sequence motifs could identify regions for rational engineering of GTs. On the other hand, sequence comparison of divergent GTs could highlight key catalytic residues.

Substrate recognition and binding

As discussed earlier, the hypervariable N3 and N5 loop regions are the main source of structural differences among Gtfs as well as between Gtfs and ScUGT51'. Also, since these loop regions dictate aglycone binding as shown in the case of GtfA and GtfD ((Mulichak *et al.* 2003; Mulichak *et al.* 2004), the corresponding sequence elements of ScUGT51' may be used as templates for diversifying the range of acceptors. However, in the case of GtfD, it was shown that determinants of substrate recognition were not exclusive to a single domain and may need to be considered before constructing chimeras.

All glycopeptide GTs that utilise NDP-donors belong to the GT-B fold family (Truman *et al*., 2009). Of these, TDP-hexoses are the typical NDP-sugar substrates for antibiotic GTs in the *Streptomyces* and *Actinomycete* group of microorganisms (Walsh *et al.* 2003). It has been suggested that the preference for TDP-sugars in Gtfs is as a result of a hydrophobic residue in place of the conserved Glu residue within the H-X_7-E motif in distantly related GT-B GTs that interacts with the ribose of the bound UDP. However, even though the structures of three Gtfs have been published, the principle of donor binding remains unclear especially due to the range of donor sugars utilised by this group of GTs (Truman *et al.* 2009).

Since ScUGT51' and related USGTs in this study show significant structural similarities similar to glycopeptide GTs, they could act as a platform for further development in the synthesis of antibiotic analogs. However, structure-function relationships may need further elucidation. For example upon DVV docking, the ScUGT51' model showed that the location of D13 was proximal to DVV to act as

the base catalyst with D103 located such that it may be involved assisting deprotonation of the acceptor. Even though ScUGT51' shares higher sequence identity with GtfB as they both utilise UDPG as NDP-donor and kinetic analysis of the D13A GtfB mutant negated the role of D13 as the potential base, mutational analysis of this residue for ScUGT51' would have to be performed to further confirm its exact role.

5. Conclusions

As described in this study, glycosylation improves bioavailability and enhances pharmacokinetic responses of sterol and steroidal drugs. In addition, due to the surge in drug-resistant pathogens, there is a growing need for the development of analogs of existing drugs to combat disease resistance. For example the development of vancomycin-resistant enterococci (VRE) has led to significant research in studying the function of Gtfs that are involved in the synthesis of the vancomycin group of antibiotics. This study reports the cloning, expression and functional analysis of yeast USGTs. As the mechanisms of GTs involved in glycan synthesis in *S. cerevisiae* are well characterized, putative as well as novel USGT sequences were identified based on the previously characterised, *N*-terminally truncated ScUGT51'. Following this, gene fragments corresponding to USGT sequences lacking the *N*-terminal PH and GRAM domains were cloned and consequently, the earlier reported problem of low expression levels (Vogt and Jones 2000) of these enzymes was overcome and GT expression levels were found to be further enhanced by the *E. coli* BL21 CP (DE3) strains. *N*-terminally His_6-tagged proteins were purified to appreciable yields and results from *in vitro* assays confirm that the segment of the *N*-terminal domain of USGTs comprising the PH and GRAM domains is not required for the catalytic activity of these enzymes, including the previously characterized non-truncated PpUGT51'. In addition, the promiscuity of all USGTs was assessed against a panel of sterol and steroidal acceptors. Following this, a putative USGT from *P. angusta* and novel USGT from *K. lactis* were identified and characterized albeit with differential regiospecificity. This therefore supports the view that although sequence alignments have been commonly used in the case of GTs to propose or further

highlight catalytic domains within the same GT family, they do not necessarily predict substrate specificity.

Furthermore, according to the CAZy classification of GTs, ScUGT51' and PpUGT51B1 (PpUGT51' in this study) are recognized as 3-β-sterol GTs although, to the best of our knowledge, this study for the first time reports that C-17-OH sterols can also be included within the panel of acceptors, although with a significantly lower affinity compared with C-3 –OH sterols. This significant finding paves the way for further exploiting the existing promiscuity of this group of GTs.

Interestingly, ScUGT51' and homologs in this study showed significant similarity to structures of Gtfs. Homology modeling based on structures of glycopeptide GTs revealed the role of conserved residues of UGTs in the present study that may be involved in substrate binding and glycosyl transfer. In addition, comparison with structures of plant GTs highlighted common residues such as D103 in ScUGT51' and UGT homologs in this study that are involved in catalysis but are not present in Gtfs suggesting an evolutionary relationship between members of the GT-B superfamily. Random mutagenesis followed by screening a relatively large library of variants via UDP*G* assays revealed that substitutions within either *N*- or *C*-terminal domains resulted in improved catalytic activity. Incidentally both the double mutants displaying the highest observed USGT activity (PaEZC62 and PpEZK15) compared to the native counterparts had substitutions within the long helical tail within the *C*-terminus and therefore beneficial mutations may occur away from the catalytic site.

The scope of this study limited the use of a wide variety of NDP-sugars. As it has been established that the GTs discussed here are able to utilise acceptors other than 3-OH compounds, studies could be focussed on rational engineering (in the USS sugar recognising motif) of these GTs to utilise a diverse range of NDP-sugars such as UDP-GlcNAc and TDP sugars as shown in the case of Gtfs.

Further Work

As discussed in Chapter 4, the ScUGT51' model suggests probable residues crucial for UGT activity based on previously solved GT-B structures. Mutational analysis of these specific residues would be required to further confirm their suggested roles.

A major bottleneck for GT function elucidation has been the lack of crystal structures which in turn depends on the availability of pure enzymes. This study reports the synthesis of pure GTs which could be used for crystallisation studies of this class of pharmaceutically relevant enzymes.

Also, the lack of availability of appropriate HTS systems is currently curtailing the DE of GTs as novel biocatalysts. Rapid and sensitive screening systems are therefore required that can be tailored to assess enhanced activity and/or promiscuity.

References

Aharoni, A., K. Thieme, C. P. C. Chiu, S. Buchini, L. L. Lairson, H. M. Chen, N. C. J. Strynadka, W. W. Wakarchuk and S. G. Withers. 2006. "High-throughput screening methodology for the directed evolution of glycosyltransferases." Nature Methods 3(8):609-614.

Ahn, B. Chan, B.G. Kim, Y.M. Jeon, E.J. Lee, Y. Lim and J.H. Ahn. 2009. "Formation of Flavone Di-O-Glucosides Using a Glycosyltransferase from *Bacillus cereus*." Journal of Microbiology Biotechnology 19(4):387-390.

Alderwick, L.J., M. Seidel, H. Sahm, G.S. Besra and L. Eggeling. 2006. "Identification of a Novel Arabinofuranosyltransferase (AftA) involved in cell wall arabinan biosynthesis in *Mycobacterium tuberculosis*." The Journal of Biological Chemistry 281:15653-15661.

Andrews, S.R., E.J. Taylor, G. Pell, F. Vincent, M.-A. Ducross, G. J. Davies, J.H. Lakey and H.J. Gilbert. 2004. "The use of forced protein evolution to inverstigate and improve stability of Family 10 Xylanases." The Journal of Biological Chemistry 279(52):54369-54379.

Banerjee, A., P. Kaul, R. Sharma and U.C. Banerjee. 2003. "A high-throughput amenable colorimetric assay for enantioselective screening of nitrilase-producing microorganisms using pH sensitive indicators." Journal of Biomolecular Screening 8(5):559-565.

Barchi, J. J. 2000. "Emerging roles of carbohydrates and glycomimetics in anticancer drug design." Current Pharmaceutical Design 6(4):485-501.

Bellu, A.R. and J.A.K.W Kiel. 2003. "Selective degradation of peroxisomes in yeasts." Microscopy Research Technique 61:161-170.

Ben-David A, G. Shoham and Y. Shoham. 2008. "A universal screening assay for glycosynthases: directed evolution of glycosynthase XynB2(E335G) suggests a general path to enhance activity." Chemistry Biology 15(6):546-551.

Berg, S., J. Starbuck, J.B. Torrelles, V.D. Vissa, D.C. Crick, D. Chatterjee and P.J. Brennan. 2005. "Roles of Conserved Proline and Glycosyltransferase Motifs of EmbC in Biosynthesis of Lipoarabinomannan." The Journal of Biological Chemistry 280:5651–5663.

Bililign, T., B.R. Griffith and J.S. Thorson. 2005. "Structure, activity, synthesis and biosynthesis of aryl-C-glycosides." Natural Product Reports 6:742-760.

Birren,B., Lander,E., Galagan,J., Nusbaum,C., Devon,K., Cuomo,C.,Kellis,M., Rasmussen,M.D., Grochow,J.A., Jaffe,D., Butler,J., Alvarez,P., Gnerre,S.,

Grabherr,M., Kleber,M., Mauceli,E., Brockman,W., MacCallum,I.A., Rounsley,S., Young,S., LaButti,K., Pushparaj,V., DeCaprio,D., Crawford,M., Koehrsen,M., Engels,R., Montgomery,P., Pearson,M., Howarth,C., Larson,L., Luoma,S., White,J., Zeng,Q., Kodira,C., Yandava,C., Alvarado,L., O'Leary,S., Reedy,J. and Heitman,J. unpublished. "Annotation of the *Pichia guilliermondii* (*Candida guilliermondii*) ATCC 6260 Genome. The Broad Institute Genome Sequencing Platform.

Bonsor, D., S. F. Butz, J. Solomons, S. Grant, I.J.S. Fairlamb, M.J. Fogg and G. Grogan. 2006. "Ligation independent cloning (LIC) as a rapid route to families of recombinant biocatalysts from sequenced prokaryotic genomes." Organic Biomolecular Chemistry 4(7):1252-1260.

Bothner, B., R. Chavez, J. Wei, C. Strupp, Q. Phung, A. Schneemann and G. Siuzdak. 2000. "Monitoring enzyme catalysis with mass spectrometry." Journal of Biological Chemistry 275(18):13455-13459.

Brazier-Hicks, M., W.A. Offen, M.C. Gershater, T.J. Revett, E-K. Lim, D.J. Bowles, G.J. Davies and R. Edwards. 2007. "Characterization and engineering of the bifunctional *N*- and *O*-glucosyltransferase involved in xenobiotic metabolism in plants." PNAS 104:20238-20243.

Breton, C., L. Šnajdrová, C. Jeanneau, J. Koca and A. Imberty. 2006. "Structures and mechanisms of glycosyltransferases." Glycobiology 16(2):29R-37R.

Brockhausen, I., B. Hu, B. Lau K. Liu, K. Lau, W. A. Szarek, L. Wang and L. Feng. 2008. "Characterization of two beta-1,3-glucosyltransferases from *Escherichia coli* serotypes O56 and O152." Journal of Bacteriology 190(14):4922-4932.

Bulter, T., M. Alcalde, V. Sieber, P. Meinhold, C. Schlachtbauer and F.H. Arnold. 2003. "Functional expression of a fungal laccase in *Saccharomyces cerevisiae* by directed evolution." Applied Environmental Microbiology 69(2):987-995.

Campbell, J.A., G. J. Davies, V. Bulone and B. Henrissat. 1997. "A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities." Biochemical Journal Letters 326:929-942.

Cantarel, B.L., P.M. Coutinho, C. Rancurel, T. Bernard, V. Lombard and B. Henrissat. 2009. "The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics." Nucleic Acids Research 37:D-233-D238.

Cao, Y. and D. J. Klionsky. 2007. "Atg26 is not involved in autophagy-related pathways in *Saccharomyces cerevisiae*." Autophagy 3(1):17-20.

Cartwright, A.M., E-K. Lim, C. Kleanthous and D.J. Bowles. 2008. "A kinetic analysis of regiosecific glucosylation by two glycosyltransferases of Arabidopsis

Thaliana: Domain swapping to introduce new activities." The Journal of Biological Chemistry 283:15724-15731.

Chapman, E. and C.H. Wong. 2002. "A pH sensitive colorimetric assay for the high-throughput screening of enzyme inhibitors and substrates: A case study using kinases." Bioorganic Medicinal Chemistry 10(3):551-555.

Charnock, S. J. and G. J. Davies. 1999. "Structure of the nucleotide-diphospho-sugar transferase, SpsA from *Bacillus subtilis*, in native and nucleotide-complexed forms." Biochemistry 38:6380-6385.

Charnock, S. J., B. Henrissat and G. J. Davies. 2001. "Three-dimensional structures of UDP-sugar glycosyltransferases illuminate the biosynthesis of plant polysaccharides." Plant Physiology 125(2):527-531.

Cohen, N., S. Abramov, Y. Dror and A. Freeman. 2001. "*In vitro* enzyme evolution: the screening challenge of isolating the one in a million." Trends in Biotechnology 19(12):507-510.

Compain, P. and O.R. Martin. 2001. "Carbohydrate mimetics-based glycosyltransferase inhibitors." Bioorganic Medicinal Chemistry 9(12):3077-3092.

Coutinho, P.M., E. Deleury, G.D. Davies and B. Henrissat. 2003. "An Evolving Hierarchical Family Classification for Glycosyltransferases." Journal of Molecular Biology 328(2):307-317.

Cowan, D. A., A. T. Kicman, C.K Garfias and H.J. Welchman. 2008. "Ion trap MS/MS of intact testosterone and epitestosterone conjugates - Adducts, fragile ions and the advantages of derivatisation." Steroids 73(6):621-628.

Dalby, P. A. 2007. "Engineering enzymes for biocatalysis." Recent Patents on Biotechnology 1(1):1-9.

Dalby, P.A. 2003. "Optimising enzyme function by directed evolution." Current Opinion in Structural Biology 13(4):500-505.

Deng, C. and R.R. Chen. 2004. "A pH-sensitive assay for galactosyltransferase." Analytical Biochemistry 330(2):219-226.

Dove, A. 2001. "The bittersweet promise of glycobiology." Nature Biotechnology 19(10):913-917.

Dujon,B., Sherman,D., Fischer,G., Durrens,P., Casaregola,S., Lafontaine,I., De Montigny,J., Marck,C., Neuveglise,C., Talla,E., Goffard,N., Frangeul,L., Aigle,M., Anthouard,V., Babour,A., Barbe,V., Barnay,S., Blanchin,S., Beckerich,J.M., Beyne,E., Bleykasten,C., Boisrame,A., Boyer,J., Cattolico,L., Confanioleri,F., De Daruvar,A., Despons,L., Fabre, E., Fairhead,C., Ferry-

Dumazet,H., Groppi,A., Hantraye,F., Hennequin,C., Jauniaux,N., Joyet,P., Kachouri,R., Kerrest,A., Koszul,R., Lemaire,M., Lesur,I., Ma,L., Muller,H., Nicaud,J.M., Nikolski,M., Oztas,S., Ozier-Kalogeropoulos,O., Pellenz,S., Potier,S.,Richard,G.F., Straub,M.L., Suleau,A., Swennen,D., Tekaia,F., Wesolowski-Louvel,M., Westhof,E., Wirth,B., Zeniou-Meyer,M.,Zivanovic,I., Bolotin-Fukuhara,M., Thierry,A., Bouchier,C.,Caudron,B., Scarpelli,C., Gaillardin,C., Weissenbach,J., Wincker,P. and Souciet,J.L. 2004. "Genome Evolution in yeasts." Nature 430, 35-44.

Emanuelsson, O., S. Brunak, G. von Heijne and H. Nielsen. 2007. "Locating proteins in the cell using TargetP, SignalP and related tools." Nature Protocols 2:953-971.

Erb, A., H. Weiss, J. Harle and A. Bechthold. 2009. "A bacterial glycosyltransferase gene toolbox: Generation and applications." Phytochemistry 70(15-16):1812-1821.

Evans, O.P. and D. R. O'reilly. 1998. "Purification and kinetic analysis of a baculovirus ecdysteroid UDP-glucosyltransferase." Biochemical Journal(3):1265-1270.

Finn, R.D., J. Tate, J. Mistry, P.C. Coggill, S.J. Sammut, H.R. Hotz, G. Ceric, K. Forslund, S.R. Eddy, E.L.L. Sonnhammer and A. Bateman. 2008. "The *Pfam* protein families database." Nucleic Acids Research 36(Database issue):D281-D288.

Fukuchi-Mizutani, M. , H. Okuhara, Y. Fukui, M. Nakao, Y. Katsumoto, K. Yonekura-Sakakibara, T. Kusumi, T. Hase and Y. Tanaka. 2003. "Biochemical and molecular characterization of a novel UDP-glucose:anthocyanin 3'-O-glucosyltransferase, a key enzyme for blue anthocyanin biosynthesis, from gentian." Plant Physiology 132(3):1652-1663.

Fulton, Z., A. McAlister, M.C.J. Wilce, R. Brammananth, L. Zaker-Tabrizi, M.A. Peruginl, S.P. Bottomley, R.L. Coppel, P.K. Crellin, J. Rossjohn and T. Beddoe. 2008. "Crystal structure of a UDP-glucose-specific glycosyltransferase from a *Mycobacterium* Species." The Journal of Biological Chemistry 283:27881-27890.

Gachon, C. M. M., M. Langlois-Meurinne and P. Saindrenan. 2005. "Plant secondary metabolism glycosyltransferases: the emerging functional analysis." Trends in Plant Science 10(11):542-549.

Gorelik, E., U. Galili and A. Raz. 2001. "On the role of cell surface carbohydrates and their binding proteins (lectins) in tumor metastasis " Cancer Metastasis Reviews 20(3-4):245-277.

Gosselin, S., M. Alhussaini, M. B. Streiff, K. Takabayashi and M. M. Palcic. 1994. "A continuous spectrophotometric assay for glycosyltransferases." Analytical Biochemistry 220(1):92-97.

Goto, M. 2007. "Protein O-glycosylation in fungi: diverse structures and multiple functions." Bioscience Biotechnology Biochemistry 71(6):1415-1427.

Grille, S., A. Zaslawski, S. Thiele, J. Plat and D. Warnecke. 2010. "The functions of steryl glycosides come to those who wait: Recent advances in plants, fungi, bacteria and animals." Progress in Lipid Research 49(3):262-288.

Ha, S., D. Walker, Y. Shi and S. Walker. 2000. "The 1.9 Å crystal structure of *Escherichia coli* MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis." Protein Science 9:1045-1052.

Hacker, C. V., C. M. Brasier and K. W. Buck. 2005. "A double-stranded RNA from a Phytophthora species is related to the plant endornaviruses and contains a putative UDP glycosyltransferase gene." Journal of general Virology 86:1561-1570.

Hahismoto, S., S Yazawa, T. Asao, A. Faried, T. Nishimura, K. Tsuboi, T. Nakagawa, T. Yamauchi, N. Koyama, K. Umehara, A.R. Saniabadi and H. Kuwano. 2008. "Novel sugar-cholestanols as anticancer agents against peritoneal dissemination of tumor cells." Glycoconjugate Journal 25:531-544.

Hancock, S.M., M. D. Vaughan and S. G. Withers. 2006. "Engineering of glycosidases and glycosyltransferases." Current Opinion in Chemical Biology 10(5):509-519.

Hans, Judith, Wolfgang Brandt and Thomas Vogt. 2004. "Site-directed mutagenesis and protein 3D-homology modelling suggest a catalytic mechanism for UDP-glucose-dependent betanidin 5-O-glucosyltransferase from Dorotheanthus bellidiformis." The Plant Journal 39(3):319-333.

Hansen, E.S., S.A. Osmani, C. Kristensen, B.L. Moller and J. Hansen. 2009. "Substrate specificities of family 1 UGTs gained by domain swapping." Phytochemistry 70(4):473-482.

Hansen, S.F., E. Bettler, A. Rinnan, S.B. Engelsen and C. Breton. 2010. "Exploring genomes for glycosyltransferases." Molecular Biosystems:1773-1781.

Hashimoto, K., T. Tokimatsu, S. Kawano, A.C. Yoshizawa, S. Okuda, S. Goto and M. Kanehisa. 2009. "Comprehensive analysis of glycosyltransferases in eukaryotic genomes for structural and functional characterization of glycans." Carbohydrate Research 344(7):881-887.

Hauss, D.J. 2007. "Oral lipid-based formulations." Advanced Drug Delivery Reviews 59:667-676.

Hefner, T., J. Arend, H. Warzecha, K. Siems and J. Sto¨ckigt. 2002. "Arbutin synthase, a novel member of the NRD1beta glycosyltransferase family, is a unique multifunctional enzyme converting various natural products and xenobiotics." Bioorganic Medicinal Chemistry 10(6):1731-1741.

Hefner, T. and S. Joachim. 2003. "Probing suggested catalytic domains of glycosyltransferases by site-directed mutagenesis." European Journal of Biochemistry 270(3):533-538.

Henrissat, B. and G. Davies. 1997. "Structural and sequence-based classification of glycoside hydrolases." Current Opinion in Structural Biology 7(5):637-644.

Hibbert, and P. A. Dalby. 2005. "Directed evolution strategies for improved enzymatic performance." Microbial Cell Factories 4(29):1-6.

Higashi, T. and K. Shimada. 2004. "Derivatization of neutral steroids to enhance their detection characteristics in liquid chromatography-mass spectrometry." Analytical and Bioanalytical Chemistry 378(4):875-882.

Hoffmeister, D., K. Ichinose and A. Bechthold. 2001. "Two sequence elements of glycosyltransferases involved in urdamycin biosynthesis are responsible for substrate specificity and enzymatic activity." Chemistry Biology 8(6):557-567.

Hoffmeister, D., B. Wilkinson, G. Foster, P.J. Sidebottom, K. Ichinose and A. Bechthold. 2002. "Engineered Urdamycin Glycosyltransferases are broadened and altered in substrate specificity." Chemistry Biology 9:287-295.

Hu, Y., L. Chen, S. Ha, B. Gross, B. Falcone, D. Walker, M. Mokhtarzadeh and S. Walker. 2003. "Crystal structure of the MurG:UDP-GlcNAc complex reveals common structural principles of a superfamily of glycosyltransferases." PNAS 100:845-849.

Hu, Y. and S. Walker. 2002. "Remarkable Structural Similarities between Diverse Glycosyltransferases." Chemistry Biology 9(12):1287-1296.

Idnurm, A., D.C. Warnecke, E. Heinz and B.J. Howlett. 2003. "Characterisation of neutral trehalase and UDP-glucose:sterol glucosyltransferase genes from the plant pathogenic fungus *Leptosphaeria maculans*." Physiological and Molecular Plant Pathology 62(5):305.

Igura, M., M. Maita, J. Kamishikiryo, M. Yamada, T. Obita, K. Maenaka and D. Kohda. 2008. "Structure-guided identification of a new catalytic motif of oligosaccharyltransferase." EMBO Journal 27:234-243.

Jackson, R. G., E. K. Lim, Y. Li, M. Kowalczyk, G. Sandberg, J. Hoggett, D.A. Ashford and D.J. Bowles. 2001. "Identification and biochemical characterization of an Arabidopsis indole-3-acetic acid glucosyltransferase." Journal of Biological Chemistry 276(6):4350-4356.

Janes, L.E., A.C. Löwendahl and R.J. Kazlauskas. 1998. "Quantitative screening of hydrolase libraries using pH indicators: identifying active and enantioselective hydrolases." Chemistry - a European Journal 4(11):2324.

Jantti, S.E., A. Kiriazis, R. R. Reinila, R. K. Kostiainen and R. A. Ketola. 2007. "Enzyme-assisted synthesis and characterization of glucuronide conjugates of neuroactive steroids." Steroids 72(3):287-296.

Johnston, M., L. Hillier, L. Riles, K. Albermann, B. André, W. Ansorge, V. Benes, M. Brückner, H. Delius, E. Dubois, A. Düsterhöft, KD. Entian, M. Floeth, A. Goffeau, U. Hebling, K. Heumann, D. Heuss-Neitzel, H. Hilbert, F. Hilger, K. Kleine, P. Kötter, E.J. Louis, F. Messenguy, H.W. Mewes and J.D. Hoheisel. 1997. "The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XII." Nature 387:87-90.

Kapitonov, D. and R. K. Yu. 1999. "Conserved domains of glycosyltransferases." Glycobiology 9(10):961-978.

Keegstra, K. and N. Raikhel. 2001. "Plant glycosyltransferases." Current Opinion in Plant Biology 4(3):219-224.

Kim, B.-G., S.H. Sung, N.R. Jung, Y. Chong and J.H. Ahn. 2010. "Biological synthesis of isorhamnetin 3-*O*-glucoside using engineered glucosyltransferase." Journal of Molecular Catalysis B: Enzymatic 63(3):194-199.

Kim, M. W., E. J. Kim, J. Y. Kim, J.S. Park, D-B. Oh, Y-I. Shimma, Y. Chiba, Y. Jigami, S.K. Rhee and H.A. Kang. 2006. "Functional characterization of the Hansenula polymorpha HOC1, OCH1, and OCR1 genes as members of the yeast OCH1 mannosyltransferase family involved in protein glycosylation." Journal of Biological Chemistry 281(10):6261-6272.

Klement, M.l, L. Ojemyr, K. E. Tagscherer, G. Widmalm and A. Wieslander. 2007. "A processive lipid glycosyltransferase in the small human pathogen *Mycoplasma pneumoniae*: involvement in host immune response." Molecular Microbiology 65(6):1444-1457.

Klionsky, D.J., J.M. Cregg, W.A. Dunn, S.D. Emr, Y. Sakai, I.V. Sandoval, A. Sibirny, S. Subramani, M. Thumm, M. Veenhuis and Y. Ohsumi. 2003. "A unified nomenclature for yeast autophagy-related genes." Developmental Cell 5:539-545.

Klutts, J.S., A. Yoneda, M.C. Reilly, I. Bose and T.L Doering. 2006. "Glycosyltransferases and their products: cryptococcal variations on fungal themes." FEMS Yeast Research 6:499-512.

Kohara, A. , C. Nakajima, S. Yoshida and T. Muranaka. 2007. "Characterization and engineering of glycosyltransferases responsible for steroid saponin biosynthesis in Solanaceous plants." Phytochemistry 68(4):478-486.

Krauth, C., M. Fedoryshyn, C. Schleberger, A. Luzhetskyy and A. Bechthold. 2009. "Engineering a function into a glycosyltransferase." Chemistry Biology 16(1):28-35.

Laemmli, U.K. 1970. "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." Nature 227:680-685.

Lairson, L.l, B. Henrissat, G. J. Davies and S. G. Withers. 2008. "Glycosyltransferases: structures, functions, and mechanisms." Annual Review of Biochemistry 77:521-555.

Lebrun, A.H., C. Wunder, J. Hildebrand, Y. Churin, U. Za¨hringer, B. Meyer Lindner, T. F. , E. Heinz and D. Warnecke. 2006. "Cloning of a cholesterol-alpha-glucosyltransferase from *Helicobacter pylori*." Journal of Biological Chemisty 281(38):27765-27772.

Lee, H. , M. Kobayashi, P. Wang, J. Nakayama, P. H. Seeberger and M. Fukuda. 2006. "Expression cloning of cholesterol alpha-glucosyltransferase, a unique enzyme that can be inhibited by natural antibiotic gastric mucin O-glycans, from Helicobacter pylori." Biochemical and Biophysical Research Communications 349(4):1235-1241.

Li, X. F., M. S. Ma, A. Cheng, J. Zheng and Y.K. Tam. 2002. "Determination of testosterone and its metabolites using liquid chromatography with elevated column temperature and flow-rate gradient." Analytica Chimica Acta 457(2):165-171.

Lim, E-K. and D.J. Bowles. 2004. "A class of plant glycosyltransferases involved in cellular homeostasis." The EMBO Journal 23:2915-2922.

Lin, C. H., C. W. Lin and K. H. Khoo. 2008. "Proteomic identification of specific glycosyltransferases functionally implicated for the biosynthesis of a targeted glyco-epitope." Proteomics 8(3):475-483.

Liu, J. and A. Mushegian. 2003. "Three monophyletic superfamilies account for the majority of the known glycosyltransferases." Protein Sci 12:1418-1431.

Loutre, C., D.P. Dixon, M. Brazier, M. Slater, D.J. Cole and R. Edwards. 2003. "Isolation of a glucosyltransferase from *Arabidopsis thaliana* active in the metabolism of the persistent pollutant 3,4-dichloroaniline." The Plant Journal 34:485-493.

Love, K.R., J.G. Swoboda, C.J. Noren and S. Walker. 2006. "Enabling Glycosyltransferase evolution: a facile substrate-attachment strategy for phage-display enzyme evolution." ChemBioChem 7(5):753-756.

Lovering, A.L., L.H. de Castro, D. Lim and N.C Strynadka. 2007. "Structural insight into the transglycosylation step of bacterial cell-wall biosynthesis." Science 315:1373-1374.

Luzhetskyy, A. and A. Bechthold. 2008. "Features and applications of bacterial glycosyltransferases: current state and prospects." Applied Microbiology and Biotechnology 80:945-952.

Luzhetskyy, A., C. Mendez, J. A. Salas and A. Bechthold. 2008. "Glycosyltransferases, important tools for drug design." Current Topics in Medicinal Chemistry 8(8):680-709.

Luzhetskyy, A., A. Vente and A. Bechthold. 2005. "Glycosyltransferases involved in the biosynthesis of biologically active natural products that contain oligosaccharides." Molecular Biosystems 1(2):117-126.

Luzhetskyy, A., H. Weiss, A. Charge, E. Welle, A. Linnenbrink, A. Vente and A. Bechthold. 2007. "A strategy for cloning glycosyltransferase genes involved in natural product biosynthesis." Applied Microbiology and Biotechnology 75(6):1367-1375.

Mackenzie, P.I., I.S. Owens, B. Burchell, K.W. Bock, A. Bairoch, A. Belanger, S.F. Gigleux, M. Green, D.W. Hum, T. Iyanagi, D. Lancet, P. Louisot, J. Magdalou, J.R. Chowdhury, J.K. Ritter, H. Schachter, T.R. Tephly, Tipton. K.F. and D.W. Nebert. 1997. "The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence." Pharmacogenetics 7(255-269).

Madina, B.R., L.K. Sharma, P. Chaturvedi, R.S. Sangwan and R. Tuli. 2007. "Purification and characterization of a novel glucosyltransferase specific to 27-hydroxy steroidal lactones from Withania somnifera and its role in stress responses." Biochimica et Biophysica Acta (BBA) - Proteins Proteomics 1774(9):1199-1207.

Maita, N., J. Nyirenda, M. Igura, J. Kamishikiryo and D. Kohda. 2010. "Comparative structural biology of Eubacterial and Archaeal oligosaccharyltransferases." The Journal of Biological Chemistry 285:4941-4950.

Mao, Z. , H. D. Shin and R. R. Chen. 2006. "Engineering the *E. coli* UDP-Glucose Synthesis Pathway for oligosaccharide synthesis." Biotechnology Progress 22(2):369.

Masada, S., Y. Kawase, M. Nagatoshi, Y. Oguchi, K. Terasaka and H. Mizukami. 2007. "An efficient chemoenzymatic production of small molecule glucosides with in situ UDP-glucose recycling." FEBS Letters 581:2562-2566.

Meech, Robyn R. and Peter I. P. I. Mackenzie. 2010. "UGT3A: novel UDP-glycosyltransferases of the UGT superfamily." Drug metabolism reviews 42(1):45-54.

Meijer, W.H., I.J. van der Klei, M. Veenhuis and J.A.K.W Kiel. 2007. "ATG Genes involved In non-selective autophagy are conserved from yeast to man, but

the selective Cvt and Pexophagy pathways also require organism-specific genes " Autophagy 3(2):106-116.

Mille, C., P. Bobrowicz, P-A. Trinel, H. Li, E. Maes, Y. Guerardel, C. Fradin, M. Martınez-Esparza, R.C. Davidson, G. Janbon, D. Poulain and S. Wildt. 2008. "Identification of a new family of genes involved in β-1,2-Mannosylation of glycans in *Pichia pastoris* and *Candida albicans*." The Journal of Biological Chemistry 283:9724-9736.

Miyazaki, K. 2002. "Creating random mutagenesis libraries by megaprimer PCR of whole plasmid." Biotechniques 33:1033-1038.

Moehs, C.P., P.V. Allen, M. Freidman and W.R. Belknap. 1997. "Cloning and expression of solanidine UDP-glucose glucosyltransferase from potato." The Plant Journal 11(2):227-236.

Moris-Varas, F., A. Shah, J. Aikens, N. P. Nadkarni, J. D. Rozzell and D. C. Demirjian. 1999. "Visualization of enzyme-catalyzed reactions using pH indicators: rapid screening of hydrolase libraries and estimation of the enantioselectivity." Bioorganic and Medicinal Chemistry 7(10):2183-2188.

Mulichak, A. M., H. C. Losey, W. Lu, Z. Wawrzak, C. T. Walsh and R.M. Garavito. 2003. "Structure of the TDP-epi-vancosaminyltransferase GtfA from the chloroeremomycin biosynthetic pathway." Proc. Natl. Acad. Sci. USA 100(16):9238-9243.

Mulichak, A.M., H.C. Losey, C.T. Walsh and R.M. Garavito. 2001. "Structure of the UDP-Glucosyltransferase GtfB that modifies the heptapeptide aglycone in the biosynthesis of vancomycin group antibiotics." Structure 9(7):547-557.

Mulichak, A.M., W. Lu, H.C. Losey, C.T. Walsh and R.M. Garavito. 2004. "Crystal Structure of vancosaminyltransferase GtfD from the vancomycin biosynthetic Pathway: interactions with acceptor and nucleotide ligands." Biochemistry 43(18):5170-5180.

Nazarko, T.Y. , J. C. Farre , A. S. Polupanov, A. A. Sibirny and S. Subramani. 2007a. "Autophagy-related pathways and specific role of sterol glucoside in yeasts." Autophagy 3(3):263-265.

Nazarko, T.Y., A. S. Polupanov, R. R. Manjithaya, S. Subramani and A. A. Sibirny. 2007b. "The requirement of sterol glucoside for pexophagy in yeast is dependent on the species and nature of peroxisome inducers." Molecular Biology of the Cell 18(1):106-118.

Newman, D.J., G.M. Cragg and K.M. Snader. 2003. "Natural products as sources of new drugs over the period 1981-2002." Journal of Natural Products 66:1022-1037.

Offen, W., C. M-F, M. Yang, E. K-L, B.J. Davis, C.A. Tarling, C.M. Ford, D.J. Bowles and G. J. Davies. 2006. "Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification." EMBO Journal 25(6):1396-1405.

Oku, M. , D. Warnecke, T. Noda, F. Muller, E. Heinz, H. Mukaiyama, N. Kato and Y. Sakai. 2003. "Peroxisome degradation requires catalytically active sterol glucosyltransferase with a GRAM domain." EMBO Journal 22(13):3231-3241.

Osmani, S.A., S. Bak and M. Lindberg. 2009. "Substrate specificity of plant UDP-dependent glycosyltransferases predicted from crystal structures and homology modeling." Phytochemistry 70(3):325-347.

Pak, J.E., P. Arnoux, S. Zhou, P. Sivarajah, M. Satkunarajah, X. Xing and J.M. Rini. 2006. "X-ray Crystal Structure of Leukocyte Type Core 2 β-1,6-N-Acetylglucosaminyltransferase-evidence for a convergence of a metal-ion dependent glycosyltransferase mechanism." The Journal of Biological Chemistry 281:26693-26701.

Palcic, M. M. and K. Sujino. 2001. "Assays for Glycosyltransferases." Trends in Glycoscience and Glycotechnology 13(72):361-370.

Palcic, M.M. 1999. "Biocatalytic synthesis of oligosaccharides." Current Opinion in Biotechnology 10(6):616-624.

Park, S.H. , H. Y. Park, J. K. Sohng , H. C. Lee, Liou K., Y. J. Yoon and B. G. Kim. 2009a. "Expanding substrate specificity of GT-B fold glycosyltransferase via domain swapping and high-throughput screening." Biotechnology Bioengineering 102(4):988-994.

Park, S.H., H.Y. Park, B.K. Cho, Y.H. Yang, J.K. Sohng, H.C. Lee, K. Liou and B.G. Kim. 2009b. "Reconstitution of Antibiotics Glycosylation by Domain Exchanged Chimeric Glycosyltransferase." Journal of Molecular Catalysis B: Enzymatic 60:29-35.

Persson, M. and M.M. Palcic. 2008. "A high-throughput pH indicator assay for screening glycosyltransferase saturation mutagenesis libraries." Analytical Biochemistry 378(1):1-7.

Pesnot, T. and G.K. Wagner. 2008. "Novel derivatives of UDP-glucose: concise synthesis and fluorescent properties." Organic and Biomolecular Chemistry 16:2884-2891.

Potocka, A. and J. Zimowski. 2008. "Metabolism of conjugated sterols in eggplant. Part 1. UDP-glucose:sterol glucosyltransferase." Acta Biochimica Polonica 55:127-134.

Qasba, P.K., B. Ramakrishnan and E. Boeggeman. 2008. "Structure and function of beta -1,4-galactosyltransferase." Current Drug Targets 9(4):292-309.

Qasba, P.K., B. Ramakrishnan and E. Boeggeman. 2005. "Substrate-induced conformational changes in glycosyltransferases." Trends in Biochemical Sciences 30(1):53-62.

Quiros, L., R.J. Carbajo, A.F. Brana and J. A. Salas. 2000. "Glycosylation of macrolide antibiotics." The Journal of Biological Chemistry 275:11713-11720.

Rabbani, S., O. Schwardt and B. Ernst. 2006. "Glycosyltransferases: An efficient tool for the enzymatic synthesis of oligosaccharides and derivatives as well as mimetics thereof." CHIMIA 60(1-2):23-27.

Rix, U., C. Fischer, L.L. Remsing and J. Rohr. 2002. "Modification of post-PKS tailoring steps through combinatorial biosynthesis." Nat. Prod. Rep, 2002 19:542-580.

Robertson, D.E. and B.A. Steer. 2004. "Recent progress in biocatalyst discovery and optimization." Current Opinion in Chemical Biology 8(2):141-149.

Rosen, M.l., M. Edman, M. Sjostrom and A. Wieslander. 2004. "Recognition of fold and sugar linkage for glycosyltransferases by multivariate sequence analysis." Journal of Biological Chemistry 279(37):38683-38692.

Rubin-Pitel, S.B. and H. Zhao. 2006. "Recent advances in biocatalysis by directed enzyme evolution." Combinatorial Chemistry High Throughput Screening 9(4):247-257.

Ruffing, A., Z. Mao and R.R. Chen. 2006. "Metabolic engineering of Agrobacterium sp. for UDP-galactose regeneration and oligosaccharide synthesis." Metabolic Engineering 8(5):465-473.

Sambrook, J. and Russell, D. 2001. "Molecular Cloning: A laboratory manual." Editions 1, 2 and 3. Cold Spring Harbor Laboratory Press. ISBN 978-087969577-4.

Scannell, D.R., Frank, A.C., Conant, G.C., Byrne, K.P., Woolfit, M. and Wolfe, K.H. 2007. "Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication". Proc. Natl. Acad. Sci. U.S.A. 104, 8397-8402.

Seibel, J., H.-J. Jordening and K. Buchholz. 2006. "Glycosylation with activated sugars using glycosyltransferases and transglycosidases." Biocatalysis and Biotransformation 24(5):311-342.

Shao, H., X. He, L. Achnine, J.W. Blount, R.A. Dixon and X. Wang. 2005. "Crystal Structures of a multifunctional triterpene/flavonoid glycosyltransferase from *Medicago truncatula*." The Plant Cell 17:3141–3154.

Simonen, M. and Palva, I. 1993. "Protein secretion in *Bacillus* species." Microbiological Reviews 57:109–137.

Singh, V., P. Somanshi, G. Rathore, D. Kapoor and B. N. Mishra. 2009. "Gene cloning, expression and homology modeling of hemolysin gene from Aeromonas hydrophila." Protein Expression and Purification 65(1):1-7.

Stasyk, O.V., T. Y. Nazarko, O. G. Stasyk, O. S. Krasovska, D. Warnecke, J. M. Nicaud, J. M. Cregg and A. A. Sibirny. 2003. "Sterol glucosyltransferases have different functional roles in *Pichia pastoris* and *Yarrowia lipolytica*." Cell Biology International 27(11):947-952.

Staver, M.J., K. Glick and D. J. Baisted. 1978. "Uridine diphosphate glucose-sterol glucosyltransferase and nucleoside diphosphatase activities in etiolated pea seedlings." Biochemical Journal 169(2):297-303.

Stevenson, D. E. 1999. "Optimisation of UDP-glucuronyl transferase-catalysed synthesis of testosterone-beta-D-glucuronide by inhibition of contaminating beta-glucuronidase." Biotechnology Techniques 13(1):17-21.

Sugai, M., N. Takakuwa, M. Ohnishi, T. Urashima and Y. Oda. 2009. "Characterization of sterol lipids in *Kluyveromyces lactis* Strain M-16 accumulating a high amount of steryl glucoside." Journal of Oleo Science 58(2):91-96.

Thibodeaux, C.J., E. Melancon Charles, C.E. and H-W. Liu. 2007. "Unusual sugar biosynthesis and natural product glycodiversification." Nature 446(7139):1008-1016.

Thorsoe, K.S, S. Bak, Olsen C. E., A. Imberty, C. Breton and B.L. Moller. 2005. "Determination of catalytic key amino acids and UDP sugar donor specificity of the cyanohydrin glycosyltransferase UGT85B1 from Sorghum bicolor. Molecular modeling substantiated by site-specific mutagenesis and biochemical analyses." Plant Physiology 139(2):664-673.

Thorson, J. S., T.J. Hosted, J. Jiang, J.B. Biggins and J. Ahlert. 2001. "Nature's carbohydrate chemists: The enzymatic glycosylation of bioactive bacterial metabolites." Current Organic Chemistry 5:139-167.

Trubetskoy, O.V. and P.M Shaw. 1999. "A fluorescent assay amenable to measuring production of beta-D-glucuronides produced from recombinant UDP-glycosyl transferase enzymes." Drug Metabolism and Disposition 27(5):555-557.

Truman, A.W., M.V.B. Dias, S. Wu, T.L Blundell, F. Huang and J.B. Spencer. 2009. "Chimeric glycosyltransferases for the generation of hybrid glycopeptides." Chemistry Biology 16:676-685.

Unligil, U.M. and J.M. Rini. 2000. "Glycosyltransferase structure and mechanism." Current Opinion in Structural Biology 10(5):510-517.

Vogt, T. and P. Jones. 2000. "Glycosyltransferases in plant natural product synthesis: characterization of a supergene family." Trends in Plant Science 5(9):380-386.

Walsh, C.T. , H.C. Losey and C.L.F Meyers. 2003. "Antibiotic glycosyltransferases." Biochemical Society Transactions 31:487-492.

Wang, X. 2009. "Structure, mechanism and engineering of plant natural product glycosyltransferases." FEBS Letters 583:3303-3309.

Warnecke, D., R. Erdmannn, A. Fahl, B. Hube, F. Muller, T. Zank, U. Zahringer and E. Heinz. 1999. "Cloning and functional expression of UGT genes encoding sterol glucosyltransferases from Saccharomyces cerevisiae, Candida albicans, Pichia pastoris, and Dictyostelium discoideum." Journal of Biological Chemistry 274(19):13048-13059.

Warnecke, D. and E. Heinz. 2010. "Glycolipid headgroup replacement: A new approach for the analysis of specific functions of glycolipids in vivo." European Journal Of Cell Biology 89(1):53-61.

Wasan, K.M. 2001. "Formulation and Physiological and biopharmaceutical issues in the development of oral lipid-based drug delivery systems." Drug Development Industrial Pharmacy 27:267-276.

Weijers, C.A.G.M., M.C.R. Franssen and G.M. Visser. 2008. "Glycosyltransferase-catalyzed synthesis of bioactive oligosaccharides." Biotechnology Advances 26(5):436-456.

Werner, S.R. and J.A. Morgan. 2009. "Expression of a *Dianthus* flavonoid glucosyltransferase in Saccharomyces cerevisiae for whole-cell biocatalysis." Journal of Biotechnology 142(3-4):233-241.

Williams, G. J. and J. S. Thorson. 2008. "A high-throughput fluorescence-based glycosyltransferase screen and its application in directed evolution." Nature Protocols 3(3):357-362.

Williams, G.J., C. Zhang and J. S. Thorson. 2007. "Expanding the promiscuity of a natural-product glycosyltransferase by directed evolution." Nature Chemical Biology 3(10):657-662.

Wimmerová, M, S.B. Engelsen, E. Bettler, C. Breton and A. Imberty. 2003. "Combining fold recognition and exploratory data analysis for searching for

glycosyltransferases in the genome of *Mycobacterium tuberculosis*." Biochimie 85:691-700.

Wohlgemuth, R. 2005. "Development, Production nd Application of Recombinant Yeast Biocatalysts in Organic Synthesis." CHIMIA 59(10):735-740.

Yang, M., M. Brazier, R. Edwards and B.J. Davis. 2005. "High-Throughput Mass-Spectrometry Monitoring for Multisubstrate Enzymes: Determining the Kinetic Parameters and Catalytic Activities of Glycosyltransferases." ChemBioChem 6(2):346.

Yang, H., J. Chen, G. Yang, X. Zhang and Y. Li. 2007. "Mutational analysis of the zinc metalloprotease EmpA of *Vibrio anguillarum*." FEMS Microbiology Letters 267:57-63.

Yao, Y., M.H. Wang, K.Y. Zhao and C.C. Wang. 1998. "Assay for enzyme activity by following the absorbance change of pH-indicators." Journal of Biochemical and Biophysical Methods 36(2-3):119-130.

Yin, Q.Y., P.W.J. de Groot, C.G. de Koster and F.M. Klis. 2007. "Mass spectrometry-based proteomics of fungal wall glycoproteins." Trends in Microbiology 16(1):20-26.

Yin, Q.Y., P.W.J. de Groot, C.G. de Koster and F.M. Klis. 2008. "Mass spectrometry-based proteomics of fungal wall glycoproteins." Trends in Microbiology 16(1):20-26.

Yu, H. , Y. Shi, H. Luo, Z. Tian, Y. Zhu and Z. Shen. 2006. "An over expression and high efficient mutation system of a cobalt-containing nitrile hydratase." Journal of Molecular Catalysis B: Enzymatic 43:80-85.

Yuan, Y., D. Barrett, Y. Zhang, D. Kahne, P. Sliz and S. Walker. 2007. "Crystal structure of a peptidoglycan glycosyltransferase suggests a model for processive glycan chain synthesis." Proc. Natl. Acad. Sci. USA 104:5348-5353.

Zea, C.J. and N. L. Pohl. 2004. "General assay for sugar nucleotidyltransferases using electrospray ionization mass spectrometry." Analytical Biochemistry 328(2):196-202.

Zhang, Y., Y. Xiang, J. L. Van Etten and M. G. Rossmann. 2007. "Structure and function of a chlorella virus-encoded glycosyltransferase." Structure 15(9):1031-1039.

Zhao, H. , K. Chockalingam and Z. Chen. 2002. "Directed evolution of enzymes and pathways for industrial biocatalysis." Current Opinion in Biotechnology 13(2):104-110.

Appendices

Appendix A.

Primary amino acid sequences of truncated proteins with corresponding ORFs and cloning region of the YSB-LIC vector

A.1 ORFs and amino acid sequences of truncated proteins. The *N*-terminal histidine tag is shown (in blue). A start codon methionine (M) was included as part of the LIC-specific sequences (in red) in cases where the truncated sequence did not have methionine as the first residue.

>gi|577215|gb|AAB67475.1| Ylr189cp [*Saccharomyces cerevisiae*]

MGSSHHHHHHMIDENPHYKTSIKPNKSYKFGLLTIGSRGDVQPYIALGKGLIKEGHQVVIITHSEF
RDFVESHGIQFEEIAGNPVELMSLMVENESMNVKMLREASSKFRGWIDALLQTSWEVCNRRKFDIL
IESPSAMVGIHITEALQIPYFRAFTMPWTRTRAYPHAFIVPDQKRGGNYNYLTHVLFENVFWKGIS
GQVNKWRVETLGLGKTNLFLLQQNNVPFLYNVSPTIFPPSIDFSEWVRVTGYWFLDDKSTFKPPAE
LQEFISEARSKGKKLVYIGFGSIVVSNAKEMTEALVEAVMEADVYCILNKGWSERLGDKAAKKTEV
DLPRNILNIGNVPHDWLFPQVDAAVHHGGSGTTGASLRAGLPTVIKPFFGDQFFYAGRVEDIGVGI
ALKKLNAQTLADALKVATTNKIMKDRAGLIKKKISKEDGIKTAISAIYNELEYARSVTLSRVKTPR
KKEENVDATKLTPAETTDEGWTMI

ATGATTGATGAGAATCCGCACTATAAAACAAGCATAAAACCAAACAAGAGTTATAAGTTTGGATTA
TTGACGATTGGCTCAAGAGGGGATGTTCAACCATATATTGCACTTGGAAAGGGCTTGATAAAAGAA
GGTCACCAAGTGGTAATAATTACACATTCGGAATTTAGAGACTTTGTTGAAAGTCATGGAATCCAG
TTTGAAGAGATTGCTGGTAACCCAGTGGAGTTGATGTCCTTGATGGTGGAAAATGAATCAATGAAT
GTTAAAATGCTGAGAGAAGCTTCGAGCAAATTTAGGGGATGGATCGATGCTCTTTTACAGACATCT
TGGGAGGTTTGTAACAGAAGAAAATTTGATATTCTGATTGAATCACCCTCAGCTATGGTTGGTATT
CATATTACTGAAGCATTGCAAATTCCTTATTTTCGAGCATTTACCATGCCATGGACAAGAACAAGA
GCATATCCACATGCATTTATTGTACCAGATCAAAAAAGGGGCGGTAACTATAACTACCTGACACAT
GTCCTTTTCGAAAACGTTTTCTGGAAGGGAATTAGTGGGCAAGTAAACAAGTGGAGAGTTGAAACT
CTAGGTTTAGGGAAAACAAATCTTTTTCTTTTACAGCAGAATAATGTTCCTTTCTTATACAATGTT
TCTCCTACAATTTTTCCGCCATCAATTGATTTTAGCGAGTGGGTGAGAGTTACAGGATATTGGTTT
CTTGATGATAAGAGTACTTTCAAACCGCCAGCAGAATTACAAGAATTTATTTCTGAGGCAAGATCA
AAAGGAAAAAAATTGGTTTATATCGGGTTTGGTTCGATTGTAGTCTCTAATGCCAAAGAAATGACG
GAAGCTCTTGTCGAAGCAGTTATGGAAGCAGATGTATACTGCATTTTAAATAAGGGATGGTCAGAA
AGGTTGGGCGATAAGGCGGCAAAGAAGACAGAGGTAGATCTTCCAAGAAACATTTTAAATATCGGA
AACGTTCCCCATGATTGGTTGTTCCCACAAGTCGATGCTGCCGTTCACCATGGTGGTTCTGGAACT
ACTGGAGCTTCATTACGTGCTGGTTTACCTACTGTTATAAAACCTTTCTTTGGTGATCAGTTCTTT
TATGCAGGTAGAGTAGAAGACATTGGAGTTGGTATTGCTTTAAAGAAACTTAACGCTCAAACATTA
GCAGATGCTTTAAAAGTAGCTACGACGAACAAAATAATGAAGGATAGAGCTGGATTAATAAAGAAG
AAAATTTCCAAAGAAGATGGGATTAAAACTGCAATAAGTGCAATTTACAACGAACTTGAGTATGCA
AGAAGCGTTACTTTATCAAGAGTGAAAACTCCGCGCAAGAAAGAGGAAAATGTAGATGCAACCAAA
CTGACACCTGCTGAAACAACGGATGAAGGGTGGACGATGATTTAA

(1431 bp)

>gi|4768597|gb|AAD29570.1|AF091397_1 UDP-glucose: sterol glucosyltransferase [*Pichia pastoris*]

MGSSHHHHHHMNQYSTSEIRSKRRYKFVLLTIGSRGDVQPYISLAKGLLAENHKVKIVTHEEFKPW
VESYGIEFATIAGNPAELMSLMVTHKSLSVGFLKEAKEKFTGWIGELLQSSWDACQDADVLIESPS
AMAGIHIAEKLQIPYFRAFTMPWTRTRAYPHAFVVPEQKRGGSYNYLTHIIFENVFWKGISGEVNK
WREQVLMLPKTNLERLEQNKVPFLYNVSPTVFPPSMDFPHWVKVVGYWFLDEGEADSYDPPKPLLE
FMEKAKTDGKKLVYIGFGSIVVSDPKQLTEAVIDAVLSADVRCILNKGWSDRLGKQTGVEVELPEE
IYNSGNVPHDWLFGKIDASVHHGGSGTTGATLRAGIPTIIKPFFGDQFFYANRVEDIGVGIGLRKL
NSKSLSKAIKEVTTNTRIIEKAKEIGKQIQSENGVSAAIRCLYQEMEYAKKLSKSKQKYWDNQSED
ISDDSVSGSWFEV

AACCAGTATTCTACGTCAGAAATTAGATCAAAGAGGAGGTACAAATTTGTGCTATTAACTATTGGT
TCTCGAGGAGATGTACAACCCTATATATCTTTGGCGAAAGGTTTACTTGCAGAGAATCACAAAGTC
AAGATTGTGACCCATGAAGAGTTCAAACCTTGGGTAGAAAGTTATGGTATCGAATTCGCAACAATT
GCAGGTAACCCAGCGGAGCTTATGTCATTGATGGTAACTCATAAGTCATTAAGTGTAGGGTTTTTG
AAAGAGGCTAAAGAAAAATTCACTGGCTGGATTGGAGAGTTATTGCAAAGCTCGTGGGACGCATGT
CAGGATGCTGATGTGCTTATCGAATCGCCTTCCGCAATGGCAGGTATCCATATCGCAGAGAAGTTA
CAAATTCCCTATTTCCGTGCTTTTACAATGCCATGGACACGAACTAGAGCGTATCCACATGCATTT
GTGGTCCCAGAACAGAAAAGAGGAGGTTCTTACAATTATCTCACGCATATAATATTTGAGAATGTC
TTTTGGAAAGGTATTTCAGGAGAGGTGAATAAATGGCGTGAGCAAGTCTTAATGCTTCCAAAGACC
AATTTGGAGAGATTGGAACAAAACAAGGTACCATTCCTGTACAATGTGTCTCCCACGGTTTTTCCA
CCATCTATGGACTTTCCTCATTGGGTTAAGGTAGTTGGATATTGGTTTTTGGATGAGGGGGAGGCA
GATAGTTATGATCCTCCTAAGCCATTGCTGGAGTTCATGGAGAAGGCTAAGACTGATGGTAAGAAA
TTGGTATACATTGGATTTGGATCCATTGTAGTTAGTGATCCCAAACAACTGACTGAGGCTGTAATC
GACGCGGTGCTCAGTGCTGATGTGAGGTGTATTCTAAATAAAGGCTGGTCAGATAGGTTAGGTAAG
CAGACGGGAGTGGAGGTTGAACTACCTGAAGAGATTTACAATTCAGGAAACGTTCCTCACGATTGG
TTGTTTGGCAAAATCGACGCGTCCGTTCATCATGGTGGATCTGGAACAACAGGTGCAACACTCCGC
GCTGGAATCCCCACCATCATCAAACCATTTTTTGGAGATCAATTCTTTTATGCCAACAGAGTAGAA
GATATTGGAGTGGGTATAGGGCTGAGAAAGTTAAACAGTAAGTCACTTTCCAAAGCTATCAAAGAG
GTAACTACAAACACTCGTATCATAGAAAAGGCTAAAGAAATTGGAAAGCAGATCCAGTCAGAAAAC
GGTGTCTCTGCAGCCATTCGTTGCCTCTACCAGGAGATGGAATACGCCAAGAAACTTTCCAAGAGC
AAGCAGAAATACTGGGACAACCAGAGCGAAGATATTAGCGACGACAGCGTAAGCGGATCATGGTTT
GAAGTGTAA

(1395 bp)

>gi|146413689|ref|XP_001482815.1| hypothetical protein PGUG_04770 [*Pichia guilliermondii* ATCC 6260]

MGSSHHHHHHMLEDSPLIKTEVRPSTSFKFTLLTIGSRGDVQPYLALAKGLQEEGHDVTIATHSEF
KEWIEGHNVKFREIAGNPTELMALMVRHGSMSVTFIKEASSKFRSWITELLSSSWDACQGTDILIE
SPSAMGGVHIAEALGIPYMRAFTMPWTRTRAYPHAFIVPDQKKGGSYNYLTHVMFETLFWKGISGQ
VNRWRQNELGLPRTSLLKLQQTKIPFLYNISPSIFPPSVDFPDWVKVTGYWFLNEGGSNFNPPPEL
LEFLSLAKENGKKVVYIGFGSIVVKNAKSLTKAIVEAVQEADIYCILNKGWSDRLLDSDDKAEKEK
PEIDLPPEVFNSGAVPHDWLFPRIDAAVHHGGSGTTGATLRAGLPTIIKPFFGDQFFYASRVEQMG
VGLSLKKLNSKSLTKALNTVTSDFKMIEKCRSISERINHEDGVSAALEAIYSELEYAKHLSISRRA
LGYEASGLQTPVVYEQSDESDESGYDDEDESEETETDEESTDNEDITEIENVPHLDYVQNHRKSLE
SSSEEDVAIATTTTND

TTAGAAGATTCACCACTCATCAAGACCGAAGTCAGACCTTCTACTTCTTTTAAATTCACACTTTTG
ACGATTGGATCGAGAGGTGATGTTCAGCCGTATTTGGCTCTAGCGAAGGGTCTTCAAGAAGAAGGA
CACGATGTAACAATTGCAACTCACTCAGAATTCAAAGAATGGATTGAGGGACACAATGTGAAATTT
CGTGAAATTGCTGGTAATCCGACAGAACTTATGGCTTTAATGGTAAGACACGGTTCTATGTCTGTG
ACGTTTATCAAAGAGGCTAGTTCCAAGTTTCGCTCTTGGATAACGGAGCTTTTGTCTTCTAGTTGG
GACGCTTGTCAAGGAACCGACATCTTGATTGAAAGTCCAAGCGCTATGGGAGGAGTTCATATTGCT
GAAGCCCTTGGAATTCCATATATGAGAGCTTTCACAATGCCATGGACAAGAACGAGAGCCTATCCT
CACGCATTTATTGTTCCAGACCAAAAGAAGGGTGGCTCTTACAACTACTTGACTCACGTTATGTTC
GAAACCCTTTTTTGGAAAGGGATATCTGGTCAAGTGAACAGATGGCGCCAGAATGAGTTGGGATTA
CCACGCACAAGTTTACTCAAATTACAGCAAACCAAGATCCCGTTCCTATACAATATCTCCCCCTCG
ATATTTCCTCCATCTGTCGATTTCCCAGATTGGGTCAAAGTGACAGGATATTGGTTCTTGAATGAG
GGAGGTAGTAACTTTAATCCTCCTCCAGAGCTTCTAGAGTTTTTGAGTCTCGCCAAGGAAAACGGA
AAGAAGGTGGTGTATATTGGATTTGGATCTATTGTGGTGAAGAACGCAAAGAGTTTGACTAAGGCT
ATTGTGGAAGCGGTTCAAGAGGCTGACATCTACTGTATTTTGAACAAAGGATGGTCTGATCGACTT
CTGGATTCTGATGATAAAGCTGAAAAAGAGAAACCTGAGATTGACTTACCCCCAGAAGTATTCAAT
TCTGGAGCGGTTCCCCATGACTGGCTCTTCCCGCGTATCGATGCGGCCGTCCATCATGGAGGCTCG
GGTACCACGGGCGCAACACTTCGAGCTGGGCTTCCCACAATTATCAAGCCGTTCTTTGGAGATCAG
TTCTTCTATGCTTCTCGAGTGGAACAAATGGGAGTAGGGCTTTCCTTGAAAAAGCTCAATTCTAAA
TCACTCACAAAAGCATTGAACACAGTGACTTCTGATTTCAAAATGATCGAAAAGTGTCGAAGTATC
AGCGAAAGAATCAACCATGAAGATGGTGTATCTGCGGCCCTTGAAGCTATCTACTCCGAATTGGAG
TATGCCAAGCATCTCAGTATCTCAAGAAGAGCTCTGGGCTACGAAGCTTCGGGCCTGCAGACTCCT
GTTGTTTACGAGCAGAGTGACGAAAGTGACGAGAGTGGATATGACGATGAAGATGAAAGCGAAGAA
ACTGAGACGGATGAGGAGAGTACAGATAACGAAGATATTACTGAGATCGAAAATGTTCCACATCTT
GACTACGTGCAAAACCATAGAAAAAGTCTGGAGAGTAGTTCAGAGGAAGATGTGGCTATAGCGACA
ACGACAACTAACGACTAA

(1602 bp)

>gi|156848169|ref|XP_001646967.1| hypothetical protein Kpol_2000p77 [*Vanderwaltozyma polyspora* DSM 70294; *Kluyveromyces polysporus*]

MGSSHHHHHHMVEDNPYFKTKIMPTKSYNFGFLTIGSRGDVQPYIALAKGLIQEGHSVTIITHREF
KSFVECHGIDFKEIAGDPTKLMSLMVEHEAINVGMLMEASSKFRGWIHDLLVTTWEACKNLKLDIL
IESPSAMAGIHISEALQIPYFRAFTMPWTRTRAYPHAFIVPDQKRGGSFNYLTHVIFENVFWRGIC
SQVNKWRVQTLGLEKTNLAQLQQNKIPFLYNISPVIFPPAIDFDEWIKVTGYWFLDESESFEPSQE
LETFISKARKLGKKLVYIGFGSIVVNNAKEMTRAVIDSVLETDIFCILNKGWSERLGKEELRYEEE
PEYPETIFLCDSIPHDWLFPKVDAAVHHGGSGTTGATLKAGTPVVIKPFFGDQFFFASRIEDIGAG
IALKKLNVSSLSNAIKKVLTDKSIKRKAVSLKKRVAKENGVTTAINCIYSELEYARSLVVKKNHKS
SNIEFIQHPNNVNDTTKTVIPLTSMV

GTTGAGGATAATCCATATTTTAAAACAAAAATAATGCCTACAAAGAGCTATAATTTTGGGTTTTTA
ACAATTGGGTCTCGTGGAGATGTTCAGCCATATATAGCCTTGGCAAAAGGATTGATACAAGAAGGA
CATAGTGTCACGATAATTACACATAGGGAATTTAAGTCATTTGTAGAGTGTCACGGCATCGATTTT
AAAGAAATTGCGGGTGATCCTACCAAGTTAATGTCATTAATGGTAGAGCATGAAGCAATTAATGTT
GGAATGCTTATGGAAGCTTCAAGTAAATTTCGGGGTTGGATTCATGATTTGTTAGTAACTACTTGG
GAAGCATGCAAAAATTTGAAGTTGGATATCCTGATTGAATCACCATCAGCAATGGCAGGAATACAT
ATTTCAGAAGCACTTCAGATACCCTATTTCCGAGCATTCACGATGCCATGGACTAGAACAAGGGCA
TATCCGCATGCATTTATTGTGCCTGATCAAAAGAGAGGCGGCAGTTTCAATTATCTGACACATGTT
ATATTTGAAAATGTTTTCTGGAGAGGAATCTGTAGTCAAGTTAATAAATGGAGAGTTCAGACTCTA
GGATTAGAAAAGACTAATTTAGCTCAACTTCAGCAAAATAAAATTCCTTTTCTTTACAATATATCT
CCAGTAATTTTCCCACCTGCTATAGACTTTGATGAATGGATCAAAGTTACTGGATATTGGTTTTTA
GATGAATCGGAATCATTTGAGCCTAGTCAAGAATTGGAGACATTTATCTCAAAGGCTCGAAAATTG
GGTAAAAAATTAGTTTATATTGGTTTTGGCTCAATTGTTGTCAATAATGCGAAAGAGATGACAAGG
GCAGTGATTGATTCAGTTTTAGAGACAGATATATTTTGTATACTCAATAAAGGTTGGTCAGAAAGG
CTGGGGAAAGAAGAATTAAGGTATGAAGAGGAGCCAGAATATCCTGAAACTATTTTCCTATGTGAT
TCTATTCCTCACGATTGGTTATTTCCTAAAGTTGACGCAGCAGTTCATCATGGTGGTTCTGGCACT
ACTGGCGCAACACTAAAGGCAGGGACACCTGTTGTAATTAAGCCATTTTTTGGTGACCAATTTTTC
TTTGCTTCAAGGATTGAAGATATTGGAGCAGGCATTGCATTGAAAAAGCTAAATGTCAGTAGTCTA
TCCAATGCCATCAAAAAAGTCCTAACTGATAAATCAATTAAAAGGAAAGCTGTTTCATTGAAAAAG
AGAGTTGCAAAAGAAAATGGTGTCACAACGGCTATAAATTGTATATATAGTGAGTTAGAGTATGCA
AGAAGTTTAGTTGTGAAAAAAAATCATAAATCTTCAAATATCGAATTTATACAGCACCCGAATAAT
GTCAATGATACGACTAAGACTGTGATTCCGCTCACCAGCATGGTCTGA

(1434 bp)

>gi|129561955|gb|ABO31066.1| Atg26p [*Pichia angusta*]

MGSSHHHHHHMIEEHPLTKTKVRPLKSYRFTLLTIGSRGDVQPYIALGKALMKEGHQVRIVTHAEF
EPWIKKHGIRFASIAGDPSELMALMVTHPTINYNFIKEAKSKFRSWIDDLLVTSWKACQDTDILIE
SPSSICGIHIAEKLQIPYFRAFTMPWTRTRAYPHAFMVPDQKLGGAYNYMTHVAFENGYWRGTAHQ
VNKWRVETLGLPKTSLAEMKQNNVPFLYNVSPTVFPPSVDFAEWVKVTGYWFLDESETYQPPEVLT
KFIEQARKDGKKVVYIGFGSIVVSKPSELTQAVVDAVLEADVRCILNKGWSDRLGTKTEIEVVLPP
EIYNAGSVPHDWLFPQIDAAVHHGGSGTTGASLRFGVPTIIKPFFGDQKFYAGRVEDLGCGVSLKD
LNYKSLARALKEVTTNTRIIEKAKLVGARIRSETGVQTAIETIYNEMEYARSLSISKVKQVSVVKS
DEEFDDDKDEEVEGSWLLV

ATAGAAGAGCATCCATTGACCAAAACCAAAGTTAGGCCGTTGAAATCTTACAGATTCACGCTGTTG
ACGATTGGATCCAGAGGAGACGTTCAGCCTTATATTGCATTGGGAAAGGCCTTGATGAAAGAGGGC
CACCAGGTTCGCATTGTCACGCATGCCGAGTTCGAGCCTTGGATCAAGAAACATGGTATTAGGTTT
GCATCGATAGCGGGAGATCCATCTGAACTCATGGCATTGATGGTCACCCATCCAACCATAAATTAC
AACTTTATAAAGGAGGCCAAGTCAAAGTTCCGGAGCTGGATCGATGACCTGTTGGTCACATCATGG
AAAGCCTGTCAGGACACAGATATTCTGATCGAGTCTCCGTCGTCAATATGTGGTATTCACATTGCT
GAGAAGCTGCAAATTCCTTACTTCCGTGCTTTTACCATGCCGTGGACACGTACGAGAGCGTATCCT
CACGCCTTCATGGTTCCCGATCAGAAACTGGGCGGCGCATACAACTATATGACCCATGTGGCTTTC
GAAAATGGATACTGGCGAGGAACCGCTCATCAGGTGAATAAATGGCGTGTTGAAACGTTAGGGCTG
CCTAAAACGTCTCTAGCAGAGATGAAACAGAATAACGTTCCTTTCCTTTATAACGTTTCGCCAACC
GTCTTTCCTCCTTCTGTGGATTTTGCTGAGTGGGTCAAAGTTACCGGTTACTGGTTTTTGGATGAA
AGCGAGACGTACCAGCCACCCGAGGTTCTGACTAAGTTCATCGAGCAGGCGAGAAAAGACGGGAAA
AAGGTGGTTTACATTGGATTTGGGTCCATCGTCGTTTCCAAGCCTTCAGAACTCACCCAAGCGGTT
GTGGACGCTGTTCTTGAGGCCGATGTCAGGTGCATTCTGAATAAAGGGTGGTCAGATCGTCTGGGA
ACGAAGACCGAAATTGAGGTGGTGCTTCCACCTGAAATATATAATGCTGGAAGCGTGCCCCATGAT
TGGCTATTTCCTCAGATCGATGCTGCTGTTCATCACGGAGGCTCTGGAACGACAGGAGCTTCGTTG
CGGTTCGGGGTGCCGACAATCATCAAGCCGTTTTTTGGAGACCAGAAGTTTTATGCAGGAAGAGTC
GAAGACCTTGGATGTGGTGTTAGTCTGAAAGATTTGAACTATAAGAGTCTGGCGAGGGCTCTCAAG
GAGGTGACCACCAACACAAGGATTATCGAAAAGGCTAAACTTGTGGGAGCGAGAATCCGGTCCGAA
ACAGGCGTTCAAACTGCCATTGAAACAATTTACAACGAAATGGAGTACGCTCGGAGTTTATCAATC
AGCAAGGTGAAACAAGTTTCTGTTGTAAAGAGCGACGAGGAGTTTGATGACGATAAAGACGAGGAA
GTTGAAGGCTCGTGGTTACTGGTTTGA

(1413 bp)

>gi|50304663|ref|XP_452287.1| unnamed protein product [*Kluyveromyces lactis*]

MGSSHHHHHHMVEKNPYFTTVIKPSKSYKFGLLTIGSRGDVQPYIALAKGLQAEGHEVIILTHGEF
KDWIVSHNIGFREISGNPAELISLMVQHGSMNMGLLRDASTNFSTWISSLLDTAWEGCQGIDILIE
SPSAMAGIHIAEALRIPYFRAFTMPWTRTRAYPHAFIVPDQKRGGNYNYFTHVLFENIFWKGISGK
VNEWRETKLKLPKTNLVSMQQNRVPFLYNVSPIVFPPSVDFNEWIKVTGYWFLDEKRSYKPPAEFM
EFLNKARELKKKVVYIGFGSIVVNDPEKMTDTIIEAVRDAGVYCVLNKGWSNRFGDPLAKKIDKEL
PSYIYNSGDVPHDWLFTKIDATVHHGGSGTTGASLRAGLPTIIKPFFGDQFFYASRVEDIGAGVAL
KKLNRSSLAKALKEVTTNTRIIQKARQIGESISKEHGVATAIGAIYSELGYARSLIKTKNPVDDKE
MEAASTKLSNDAVVTAKGNEKEEYSSEGSGSNDGSWLLI

GTTGAAAAAAACCCGTATTTCACCACTGTAATCAAACCTAGTAAAAGCTACAAATTCGGATTACTT
ACAATCGGTTCAAGAGGTGATGTACAACCATATATTGCATTGGCAAAGGGTTTGCAAGCCGAAGGT
CATGAAGTTATTATACTAACTCATGGAGAATTTAAGGATTGGATAGTTTCTCATAATATAGGATTT
AGAGAAATATCTGGTAATCCTGCAGAACTAATATCTTTGATGGTACAGCATGGATCAATGAACATG
GGTTTATTGAGGGACGCTTCTACAAACTTTTCTACTTGGATCTCATCTCTGTTGGATACAGCTTGG
GAGGGATGTCAAGGTATAGATATACTAATCGAATCTCCTAGTGCCATGGCAGGGATACATATAGCA
GAAGCTTTGAGAATACCATATTTTAGAGCGTTTACTATGCCTTGGACTAGAACTAGGGCTTATCCA
CATGCATTTATAGTACCAGATCAAAAAAGAGGTGGCAACTATAACTATTTTACACATGTTCTTTTC
GAGAACATCTTTTGGAAAGGAATCAGTGGTAAGGTTAATGAATGGAGAGAAACCAAATTAAAATTG
CCCAAGACGAATCTAGTTTCGATGCAGCAAAATCGTGTTCCATTTTTATACAATGTGTCTCCGATA
GTCTTTCCACCAAGTGTCGATTTCAATGAATGGATCAAAGTAACAGGTTACTGGTTTCTTGATGAA
AAACGGTCCTACAAACCCCCTGCAGAATTTATGGAATTTCTCAACAAGGCTAGGGAATTGAAGAAA
AAAGTGGTATATATTGGATTCGGATCGATAGTAGTTAATGACCCCGAAAAAATGACAGATACTATT
ATTGAAGCTGTACGAGATGCAGGAGTTTATTGTGTTTTGAATAAAGGCTGGTCTAATAGGTTTGGA
GATCCACTGGCAAAAAAGATCGATAAAGAGTTGCCAAGTTATATATATAATTCAGGTGATGTGCCA
CACGACTGGCTATTCACAAAAATTGATGCTACAGTTCATCATGGCGGTTCGGGAACTACAGGCGCG
TCTTTAAGAGCTGGTTTGCCTACCATTATCAAACCATTCTTTGGGGATCAATTCTTTTACGCTAGC
AGAGTTGAAGATATTGGCGCTGGTGTCGCTTTGAAAAAATTGAATAGGAGCTCGTTAGCGAAAGCG
TTGAAAGAAGTTACCACTAACACGAGGATCATCCAAAAGGCCAGGCAAATAGGTGAGAGTATTTCT
AAAGAGCATGGGGTTGCTACGGCCATAGGCGCAATTTATTCTGAATTAGGATACGCTAGAAGTTTG
ATAAAAACCAAAAATCCCGTTGATGACAAAGAGATGGAGGCTGCAAGCACAAAATTATCCAACGAT
GCTGTTGTAACCGCGAAAGGAAATGAAAAAGAAGAATATAGCAGTGAAGGCTCAGGATCAAATGAT
GGATCCTGGCTTCTGATATAA

(1473 bp)

>gi|50423499|ref|XP_460332.1| hypothetical protein DEHA0E25168g [*Debaryomyces hansenii* CBS767]

MGSSHHHHHHMLEDSPFYKTEIRPSTSFNFTLLTIGSRGDVQPYIALGKGLLNEGHNVTIATHSDF
EEWIVGHGIKFKTIAGNPVELMSLMVTHGSMSLSFLKEASSKFRGWIQELLDTSWKACQGSDILIE
SPSAMVGAHIAEALGIPYIRAFTMPWTRTRAYPHAFIVPDKKKGGSYNYITHLMFETVLWKGISSQ
VNKWRRELLGLPRTNLYRLAQYDIPFLYNISPTIFPPSVDFPDWVKVTGYWFLDEGAADDFEPLKE
LVEFMNKARADDKKVVYIGFGSIVVEDAKSLTKAIVEAVLNADVRCILNKGWSDRNSSPAKDNAEP
EVELPEEIYNSGSIPHDWLFPKIDAAVHHGGSGTTGATMRAGIPTIIKPFFGDQFFYSSRIEDIGA
GIGLKKLNARSLCTALKTATSDAKMITKAKKISERLKQENGVLNAIEAIYYELEYARSLILAKQHE
NTKHDLKSGTQTPVVNETNEYFDSDTYDADHDSDKESDHDQTYEQDNHSDYDVANDDNMTEIVEPS
LEDGNDTVRIAPDSGNDNTTVTDANK

TTAGAAGATTCTCCATTCTATAAGACGGAAATACGGCCTTCTACTTCCTTTAATTTTACATTATTA
ACTATTGGATCACGAGGAGACGTACAACCTTACATAGCATTAGGCAAAGGGCTATTGAATGAGGGT
CATAATGTAACAATTGCCACACACTCTGACTTCGAAGAATGGATTGTAGGACATGGCATTAAATTT
AAGACAATTGCAGGAAATCCAGTTGAATTGATGTCATTGATGGTAACCCACGGGTCAATGTCTCTC
TCGTTTCTTAAAGAAGCAAGCTCCAAATTCAGAGGTTGGATTCAGGAATTATTAGACACTAGTTGG
AAGGCTTGTCAAGGGAGCGATATTTTGATTGAGAGCCCATCAGCCATGGTGGGTGCGCATATTGCA
GAAGCATTAGGTATTCCTTACATAAGGGCGTTTACAATGCCATGGACTAGGACCAGGGCATACCCA
CATGCGTTTATTGTACCCGATAAAAAGAAGGGTGGCTCGTATAATTATATTACACATTTAATGTTC
GAGACAGTTCTTTGGAAGGGAATCAGTAGTCAGGTCAATAAATGGCGTAGAGAACTGTTGGGATTA
CCAAGAACCAATTTGTACAGACTTGCACAATATGACATACCGTTTTTATACAACATCTCACCTACA
ATATTCCCACCTTCTGTTGATTTTCCAGATTGGGTCAAAGTTACTGGCTACTGGTTTCTAGACGAA
GGAGCGGCTGATGATTTTGAACCTCTGAAAGAATTGGTTGAGTTTATGAATAAGGCAAGAGCAGAC
GATAAGAAGGTTGTTTACATTGGGTTTGGTTCTATCGTGGTGGAAGATGCCAAGAGCCTTACAAAG
GCAATAGTCGAAGCTGTTCTCAATGCAGATGTTCGCTGTATCTTGAACAAAGGTTGGTCGGACAGA
AACTCTAGCCCCGCAAAGGACAATGCGGAGCCAGAAGTTGAACTACCAGAGGAAATTTACAATTCA
GGATCTATTCCACACGATTGGTTATTTCCAAAGATAGACGCGGCTGTTCATCATGGTGGATCTGGT
ACAACTGGTGCTACAATGAGAGCTGGTATTCCCACTATAATCAAACCATTTTTTGGGGATCAATTC
TTCTATTCTTCAAGGATCGAAGATATTGGCGCTGGAATTGGGTTAAAGAAGTTAAATGCTCGTTCA
TTGTGCACTGCATTGAAAACTGCAACATCGGATGCAAAAATGATTACAAAGGCAAAGAAAATTAGC
GAAAGATTGAAACAAGAGAATGGTGTTTTGAATGCAATTGAAGCCATATATTATGAACTAGAATAT
GCAAGAAGTTTAATACTTGCTAAACAACATGAAAATACCAAGCATGACCTCAAGTCCGGTACCCAA
ACACCTGTTGTTAATGAAACAAATGAATATTTCGATAGTGACACATATGACGCAGACCACGATTCA
GATAAGGAATCAGACCACGATCAGACATATGAACAGGATAATCATAGTGATTATGATGTTGCTAAT
GACGACAACATGACCGAAATTGTAGAACCTAGTCTTGAGGATGGCAATGATACTGTAAGAATCGCA
CCTGATAGTGGAAATGATAACACTACCGTTACAGATGCTAACAAATAA

(1632 bp)

A.2 Map and cloning region of the pET-YSBLIC vector

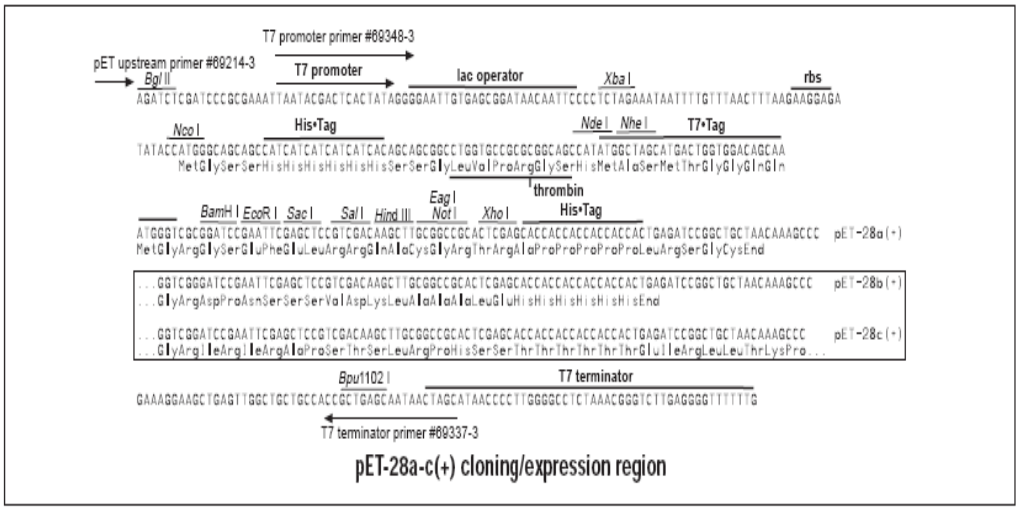


Fig. A.1 Cloning region for the pET-28a derived LIC vector *(Novagen, 2007). Target gene fragments were cloned with the inclusion of the stop codon to allow expression of an N-terminally his-tagged protein.*

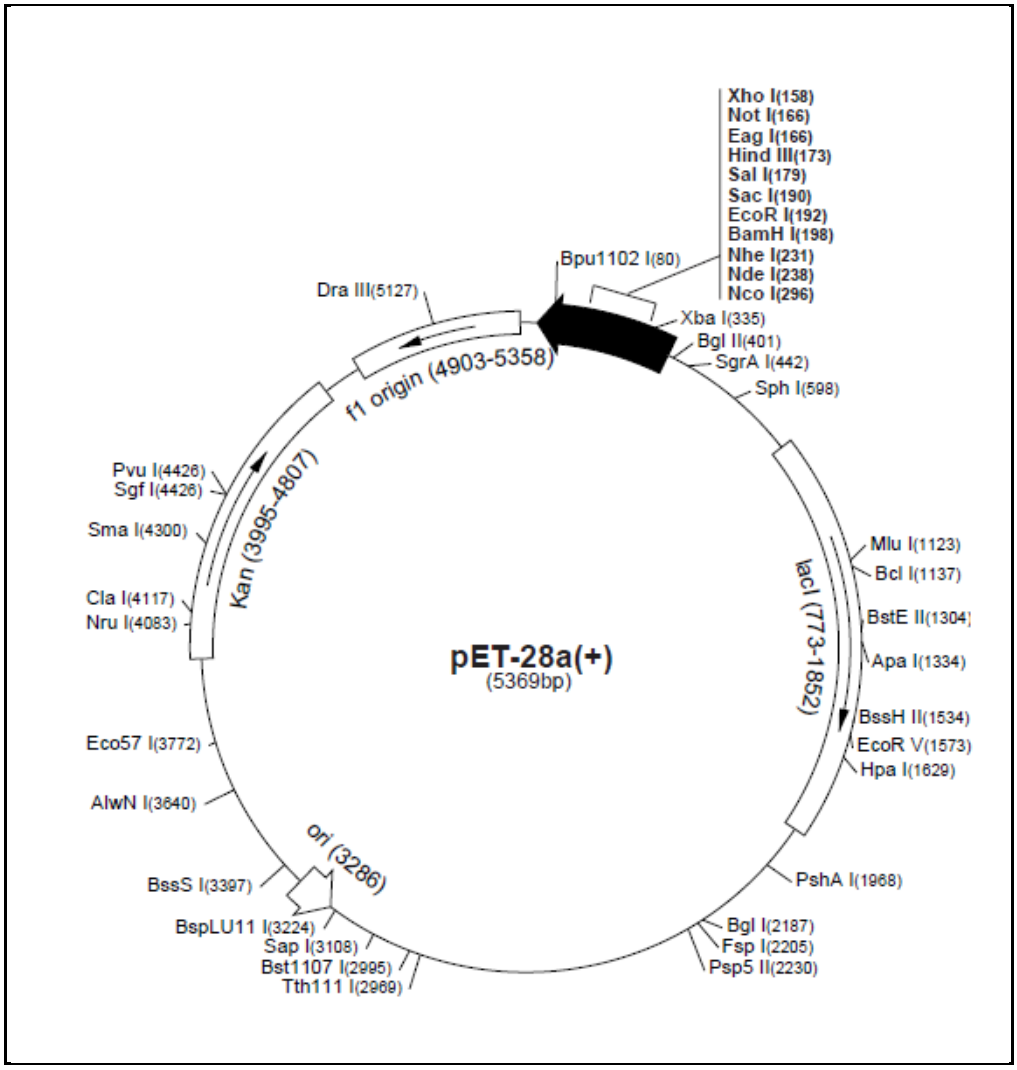
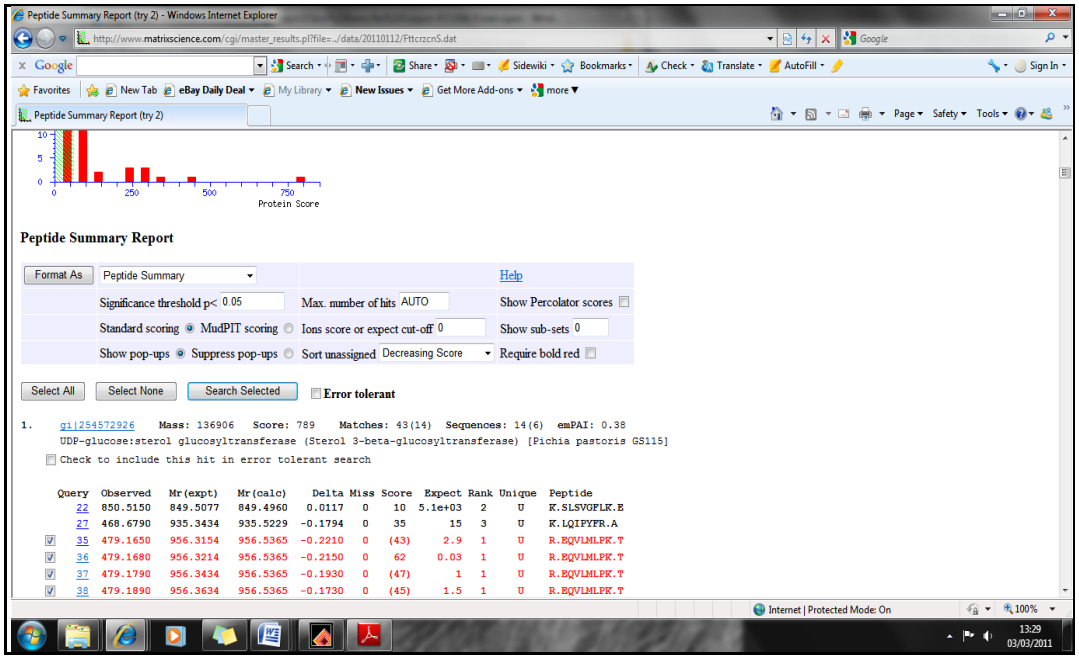
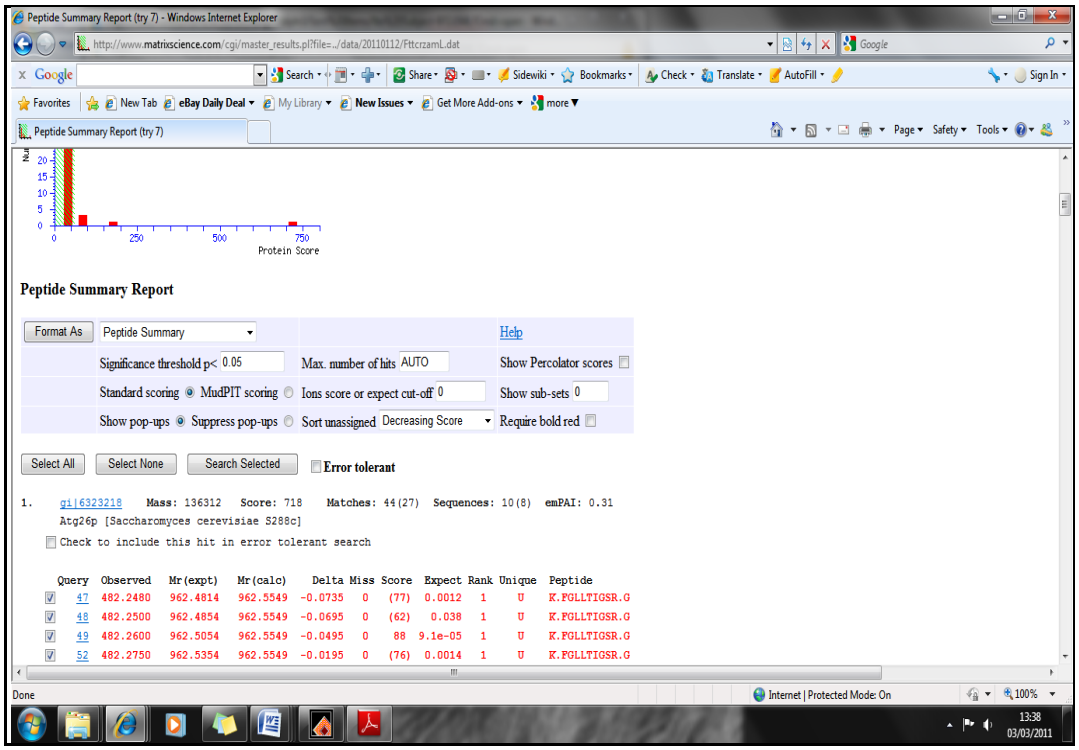
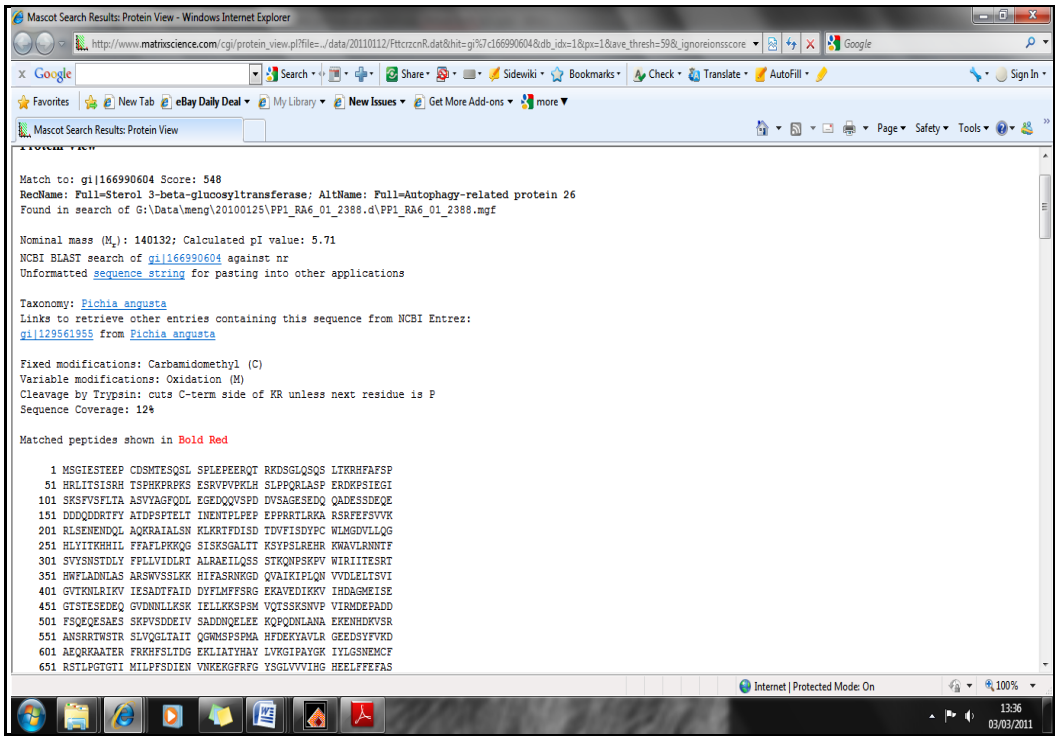


Fig. A.2 Map of pET-28a vector

Appendix B. Mascot peptide analysis outputs






Mascot Search Results: Protein View - Windows Internet Explorer
Mascot Search Results: Protein View
Match to: gi|166990604 Score: 548
RecName: Full=Sterol 3-beta-glucosyltransferase; AltName: Full=Autophagy-related protein 26
Found in search of G:\Data\meng\20100125\FP1_RA6_01_2388.d\FP1_RA6_01_2388.mgf
Nominal mass (Mr): 140132; Calculated pI value: 5.71
NCBI BLAST search of gi|166990604 against nr
Unformatted sequence string for pasting into other applications
Taxonomy: Pichia angusta
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|129561955 from Pichia angusta
Fixed modifications: Carbamidomethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: 12%
Matched peptides shown in Bold Red
1 MSGIESTEEP CDSMTESQSL SPLEPEERQT RKDSGLQSQS LTKRHFAFSP
51 HRLITSISRH TSPHKPRPKS ESRVPVPKLH SLPPQRLASP ERDKPSIEGI
101 SKSFVSFLTA ASVYAGFQDL EGEDQQVSPD DVSAGESEDQ QADESSDEQE
151 DDDQDDRTFY ATDPSPTELT INENTPLPEP EPPRRTLRKA RSRFEFSVVK
201 RLSENENDQL AQKRAIALSN KLKRTFDISD TDVFISDYPC WLMGDVLLQG
251 HLYITKHHIL FFAFLPKKQG SISKSGALTT KSYPSLREHR KWAVLRNNTF
301 SVYSNSTDLY FPLLVIDLRT ALRAEILQSS STKQNPSKPV WIRIITESRT
351 HWFLADNLAS ARSWVSSLKK HIFASRNKGD QVAIKIPLQN VVDLELTSVI
401 GVTKNLRIKV IESADTFAID DYFLMFFSRG EKAVEDIKKV IHDAGMEISE
451 GTSTESEDEQ GVDNNLLKSK IELLKKSPSM VQTSSKSNVP VIRMDEPADD
501 FSQEQESAES SKPVSDDEIV SADDNQELEE KQPQDNLANA EKENHDKVSR
551 ANSRRTWSTR SLVQGLTAIT QGWMSPSPMA HFDEKYAVLR GEEDSYFVKD
601 AEQRKAATER FRKHFSLTDG EKLIATYHAY LVKGIPAYGK IYLGSNEMCF
651 RSTLPGTGTI MILPFSDIEN VNKEKGFRFG YSGLVVVIHG HEELFFEFAS
Internet | Protected Mode: On
100%
13:36
03/03/2011

Appendix C. Protein expression using auto-induction media

Starter cultures for auto-induction media

Component	Volume/ 10 ml
Water	9.55 ml
1M $MgSO_4$	20 µl
1000 x metals	2 µl
40% (w/v) glucose	125 µl
25% (w/v) aspartate	100 µl
50 x M	200 µl
kanamycin (10 mg/ml)	100 µl

ZYM-5052 auto-induction media

Component	Volume (ml)/ 500 ml
ZY media	478.5
1M $MgSO_4$	1
1000 x metals	0.5
50 x 5052 media	10
50 x M media	10
kanamycin (10 mg/ml)	5

Stock solutions for auto-induction media

50 x M media

Component	Amount (g)/100 ml
Na_2SO_4 (anhydrous)	3.6
NH_4Cl (anhydrous)	13.4
KH_2PO_4 (anhydrous)	17.0
Na_2HPO_4 (anhydrous)	17.7

The salts were sequentially dissolved in distilled water and autoclaved prior to use.

1000 x metals solution

Component	Volume (ml)/100 ml
sterile 18.2 MΩ/cm H_2O	36
0.1 M $FeCl_3 \cdot 6H_2O$	50
(Dissolved in 100-fold dilution of 0.12 M HCl and filter sterilised using 0.2 µm filters). The following constituents were then added	
1 M $CaCl_2$ (anhydrous)	2
1 M ZnSO4·7H2O	1
0.2 M $CoCl_2 \cdot 6H_2O$	1
0.1 M $CuCl_2 \cdot 2H_2O$	2
0.2 M $NiCl_2 \cdot 6H_2O$	1
0.1 M $Na_2MoO_4 \cdot 5H_2O$	2
0.1 M $Na_2SeO_3 \cdot 5H_2O$	2
0.1 M H_3BO_3 (anhydrous)	2

All the solutions were made separately and autoclaved unless stated otherwise.

20 ml of a 25% (w/v) aspartate solution was prepared and the pH was adjusted to 7.0 by neutralisation with NaOH and autoclaved separately.

50 x 5052 media

The following constituents were sequentially dissolved in distilled H_2O

Component	Amount (g)/100 ml
Glycerol (100%)	25 g
Glucose	2.5 g
α-lactose	10 g

ZY media

10 g Tryptone and 5 g yeast extract was sequentially dissolved in 1 L distilled H_2O.

Appendix D. General use equipment

Sterilisation

Autoclave sterilisation was achieved at 121°C for 30 min using a bench-top Prestige Medical 2100 Classic autoclave or 121°C for 40 min in a Priorclave 200 L front loading autoclave. Filter sterilisation was performed through a 0.2 μm Sartorius, Ministart® filter unit.

Incubators

For the growth of bacteria in liquid cultures an Innova 44 orbital shaker Brunswick Scientific (BRS) was used. Growth of bacteria on solid media was performed in a static GallenKamp incubator.

pH meter

All adjustments to the pH of solutions and media were achieved using a Jenway Ion Meter 3340 calibrated with buffers at pH 4.0, 7.0, and 9.2.

Centrifugation:

MSE Microcentaur bench-top centrifuge was used to centrifuge PCR tubes. Centrifugation of volumes upto 1.5ml was achieved using a small Sigma (SIG) 1-15 bench top micro-centrifuge. Volumes above 1.5ml were centrifuged in a large 3K18C refrigerated bench top centrifuge (SIG) using the rotors and inserts appropriate to the application.

Agarose gel electrophoresis (AGE)

Electrophoresis of DNA was performed using an electrophoresis system (BR) powered by a power PAC 300 apparatus (BR).

Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) gel kit

Electrophoresis of proteins was achieved using A Mini-PROTEAN™ 3 Cell kit powered (BR) by a power PAC 300 apparatus (BR).

Gel documentation

Visualisation of agarose gels was achieved using a Gel Doc 2000 system utilising Quantity One software and equipped with a UV source (BR). SDS-PAGE gels were visualised using a GS-800 calibrated densitometer (BR).

Sonication

Cell lysis was achieved using a MSE Soniprep 150 ultra-sonication machine (Sanyo).

Spectrophotometer

Spectrophotometric measurements were made using a UV-visible Heλios α Spectronic Unicam spectrophotometer (TS).

PCR

All PCR reactions were performed using a Mastercycler® gradient thermocycler (Eppendorf).

Vortexing

Fisher Scientific (FIS) whirlmixer was used to vortex samples.

Waterbath

Grant (SUB series) incubator water bath used to incubate samples at digitally controlled temperatures.

Pippetting

Gilson pipettes P2-P5000 used to pipette samples and reagents from 0.01-5000 μl accurately.

Weighing Reagents

Mettler AT250 analytical balance used to measure micro gram quantities of chemicals and reagents.

Protein Purification

Akta Prime Plus™ 5.0 Series system, software version 3.0, was used for automated Ni^{2+} affinity chromatography using a High Performance Ni^{2+} Sepharose resin® (AB).

Sample concentrator

Techne Dri-Block® sample concentrator was used to evaporate solvents from samples using N_2 gas supply. Speed-Vac® sample concentrator (TS) was used to concentrate DNA samples and drying protein gel slices.

Plate images

Images of agar plates and 96-well plates were captured using the Nikon Coolpix® 995 3.34 mega pixels digital camera.

Freeze drying

Samples were freeze dried using an ALPHA 1-2 LD Freezer dryer (Christ)

Appendix E. SDS-PAGE gel images

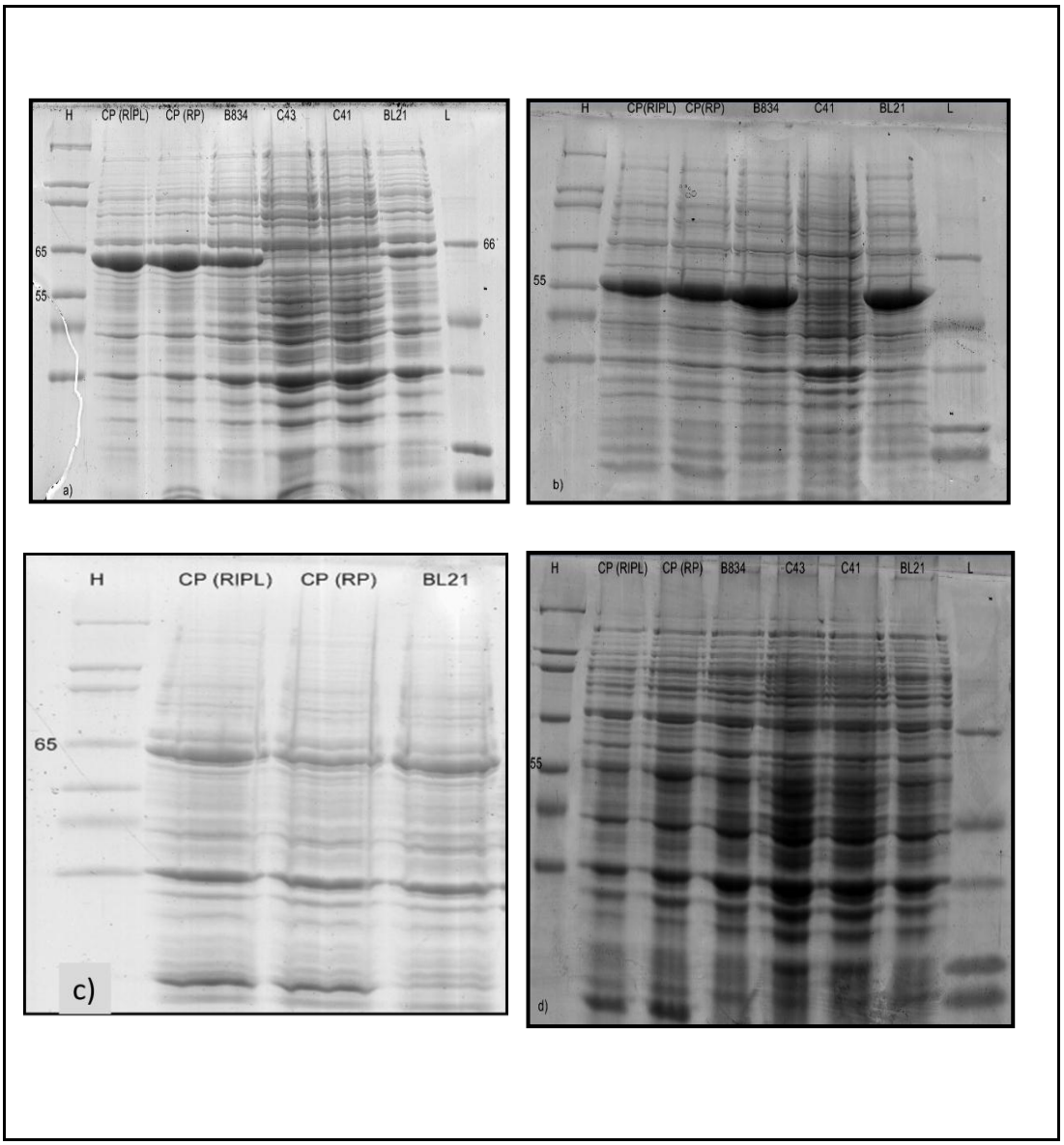


Fig E.1 SDS-PAGE analysis of membrane fractions of *E. coli* (DE3) (indicated above lanes) following expression of the Pg, Pa, Dh and Sc r-plasmids a)-d) with expected m.wt of 61, 54, 62, 55 KDa respectively. *L = protein m.wt size standard ranging between 6.5 KDa and 66 KDa. H=protein m.wt size standard ranging between 36 KDa and 205 KDa.*

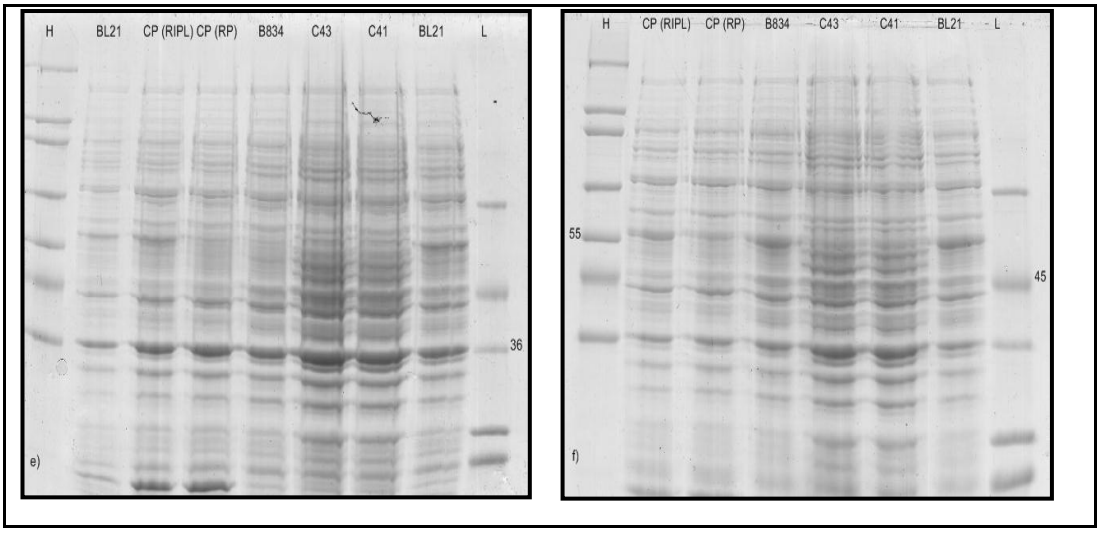


Fig. E.2 SDS-PAGE analysis of membrane fractions of *E. coli* (DE3) (indicated above lanes) following expression of the Kl and Kp r-plasmids e)-f) with expected m.wt of 56 and 55 KDa respectively. *L = protein m.wt size standard ranging between 6.5 KDa and 66 KDa. H=protein m.wt size standard ranging between 36 KDa and 205 KDa.*

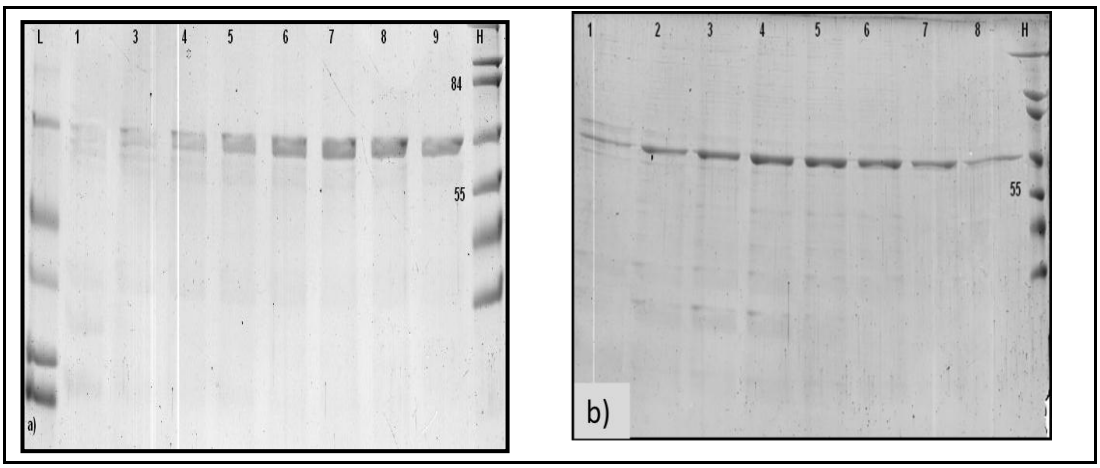


Fig. E.3 SDS-PAGE analysis of IMAC purified protein fractions. a) Protein bands of approx. 61 KDa are observed for the Pg-LIC construct transformed in CP (RP); b) Dh-LIC construct transformed in BL21 shows bands between 55 and 65 KDa (predicted m.wt: 62 KDa).

Appendix F. DNA sequencing data for mutants